

Análisis de la función de miembros de la familia Yap durante la gastrulación de teleósteos



Tesis doctoral

Javier Vázquez Marín

Director: Juan Ramón Martínez Morales



UNIVERSIDAD
**PABLO[®]
OLAVIDE**
S E V I L L A

Tesis doctoral

Análisis de la función de miembros de la familia Yap durante la gastrulación de teleósteos

*Memoria presentada por Javier Vázquez Marín para optar al título de Doctor y
dirigida por el Dr. Juan Ramón Martínez Morales*

Sevilla, octubre de 2018



U N I V E R S I D A D
**PABLO_D
OLAVIDE**
S E V I L L A

Programa de Doctorado en Biotecnología, Ingeniería y Tecnología Química (RD: 99/2011)

Financiación

Esta Tesis doctoral se ha llevado a cabo en el Centro Andaluz de Biología del Desarrollo (CABD/CSIC/UPO) y ha sido financiada mediante dos ayudas: una **Ayuda Puente-Predoctoral** convocada a concurso público por la Universidad Pablo de Olavide (Rfª.: PPI1304) y, posteriormente, por una **Ayuda para contratos predoctorales para la formación de Doctores** (antigua beca FPI), otorgada por el Ministerio de Economía y Competitividad, con referencia BES-2015-071580, y asignada al proyecto **BFU2014-53765-P (Control genético de la geometría celular en epitelios: la morfogénesis de la copa óptica de los vertebrados como paradigma experimental)**, cuyo Investigador Principal es el Dr. Juan Ramón Martínez Morales.

Asimismo, se han realizado dos estancias breves en el Centre for Organismal Studies (COS) de Heidelberg (Alemania) entre septiembre y noviembre de 2015, y entre marzo y abril de 2016. Ambas estancias han sido financiadas por medio de una beca **Travelling Fellowship** (Rfª.: DEVTF-150408), otorgada por *The Company of Biologists Limited*, y de una beca **EMBO Short Term Fellowship** (Rfª.: ASTF 369-2015), otorgada por la *European Molecular Biology Organization (EMBO)*.



Centro
Andaluz
de Biología
del Desarrollo



Centre for
Organismal
Studies
Heidelberg



“Nulla dies sine linea” (“Ni un día sin una línea”)

Cita de Plinio el Viejo atribuida a Apeles, pintor griego clásico del siglo IV a. C.

A mi familia

RESUMEN / ABSTRACT	15
INTRODUCCIÓN.....	19
1. La familia Yap/Taz	21
1.1. Introducción.....	21
1.2. Estructura.....	21
1.3. Regulación de la actividad de Yap y Taz.....	22
1.4. Yap y Taz a lo largo de la Evolución	25
2. Gastrulación en peces teleósteos	26
2.1. Introducción	26
2.2. Principales movimientos de gastrulación	27
2.3. Rutas de señalización implicadas en la gastrulación	30
2.4. Mecanismos de adhesión implicados en la gastrulación.....	33
3. Conexiones entre la actividad de la familia Yap/Taz y mecanismos implicados en la gastrulación de teleósteos	36
OBJETIVOS	39
MATERIALES Y MÉTODOS	43
1. Mantenimiento de líneas de peces teleósteos	45
2. Análisis filogenético de genes de la familia Yap.....	45
3. Generación de líneas de mutantes en pez medaka	46
4. iDamID-seq.....	48
5. RNA-seq.....	52
6. Hibridaciones in situ.....	53
7. Análisis de expresión de genes de la familia Yap	55
8. Ensayos funcionales	56
9. Análisis de trayectorias celulares durante la gastrulación.....	57
10. Análisis del grado de apoptosis y de proliferación celular en embriones mutantes.....	59

RESULTADOS	61
1. Identificación de Yap1b como nuevo miembro de la familia de proteínas Yap	63
2. Análisis de expresión de <i>yap1</i> y <i>yap1b</i> durante el desarrollo temprano.....	66
3. Identificación del conjunto de sitios de unión a la cromatina de proteínas de la familia Yap durante la embriogénesis temprana en teleósteos mediante iDamID-seq	68
4. La mutación del gen <i>yap1b</i> potencia los defectos en el desarrollo de embriones <i>yap1</i> ^{-/-} durante la embriogénesis de medaka	74
5. Identificación de genes regulados por <i>yap1</i> durante la gastrulación tardía de medaka.....	78
6. Yap1 cumple un papel importante en el establecimiento del eje corporal de medaka durante la gastrulación.....	80
DISCUSIÓN	89
1. Introducción	91
2. Evolución de parálogos de la familia Yap en teleósteos	91
3. Análisis de regiones de unión a DNA de proteínas de la familia Yap.....	92
4. Función cooperativa entre parálogos de la familia Yap durante el desarrollo temprano de teleósteos	94
5. Papel de Yap como agente inductor de migración celular durante la gastrulación de medaka	96
6. Yap1 dirige la migración de células hacia la línea media a través del control de la expresión de <i>ctgfa</i>	99
7. Observaciones finales	101
CONCLUSIONES / CONCLUDING REMARKS.....	103
MATERIAL ADICIONAL	109
REFERENCIAS BIBLIOGRÁFICAS	137
ANEXO	159
AGRADECIMIENTOS.....	173

Figura 1 - Estructura básica de las proteínas Yap y Taz.....	22
Figura 2 - Mecanismos de regulación de la actividad de Yap/Taz	25
Figura 3 - Principales movimientos celulares registrados durante la gastrulación de pez cebra	29
Figura 4 - Resumen de los principales mecanismos celulares implicados en la gastrulación de teleósteos	34
Figura 5 - Esquema del protocolo experimental de la técnica iDamID-seq	49
Figura 6 - Extracción de muestras de RNA y DNA para RNA-seq	53
Figura 7 - Procesamiento de imágenes de hibridaciones in situ (ISH)	54
Figura 8 - Esquema de diseño de morfolidos	56
Figura 9 - Análisis filogenético de proteínas de la familia Yap en teleósteos	64
Figura 10 - Análisis de expresión de genes de la familia Yap	67
Figura 11 - Perfil de iDamID-seq de proteínas de la familia Yap.....	70
Figura 12 - Identificación de motivos asociados a picos de DamID-seq en medaka.....	72
Figura 13 - Generación de alelos mutantes <i>yap1b</i> mediante tecnología CRISPR-Cas9	75
Figura 14 - Niveles de apoptosis y proliferación celular en embriones mutantes <i>yap1^{-/-}</i> <i>yap1b^{-/-}</i>	77
Figura 15 - Identificación de dianas transcripcionales de <i>yap1</i> mediante RNA-seq	79
Figura 16 - Longitud y anchura de embriones <i>yap1^{-/-}</i> de medaka	81
Figura 17 - Patrón de expresión de <i>ntl</i> en embriones <i>yap1^{+/+}</i> y <i>yap1^{-/-}</i> en embriones de medaka de estadio 16	82
Figura 18 - Inyección conjunta de morfolidos contra <i>ctgfa</i> y <i>cyr61</i> en medaka	83
Figura 19 - Obtención de embriones CRISPRantes de medaka para el gen <i>ctgfa</i>	85
Figura 20 - Películas generales de gastrulación en medaka.....	86
Figura 21 - Rastreo de trayectorias celulares en la región posterior de embriones de medaka durante la gastrulación.....	88

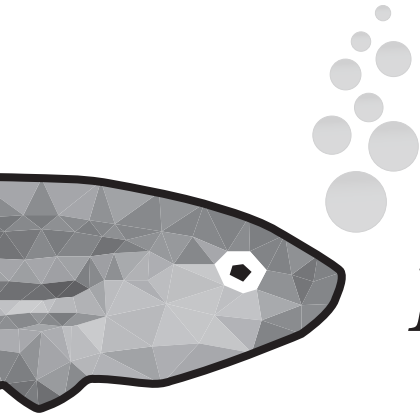
Figura Sup. 1 - Etapas de desarrollo embrionario de pez cebra	111
Figura Sup. 2 - Etapas de desarrollo embrionario de pez medaka	112
Figura Sup. 3 - Generación de alelos mutantes en el gen <i>yap1</i> de medaka mediante tecnología CRISPR-Cas9	113
Figura Sup. 4 - Análisis sinténico de parálogos de la familia Yap.....	114
Figura Sup. 5 - Análisis de la estructura del dominio de activación transcripcional de Yap1b.....	115
Figura Sup. 6 - Historia evolutiva de la familia Yap en teleósteos	116
Figura Sup. 7 - Correlación entre réplicas en experimentos de iDamID-seq	117
Figura Sup. 8 - Perfiles iDamID-seq adicionales de <i>yap1</i> y <i>yap1b</i> en medaka.....	118
Figura Sup. 9 - Solapamiento de picos y distancia al TSS más cercano	119
Figura Sup. 10 - Identificación de motivos asociados a picos de iDamID-seq en pez cebra	120
Figura Sup. 11 - Niveles de apoptosis en embriones de medaka con al menos una copia WT de <i>yap1</i>	120
Figura Sup. 12 - Análisis fenotípico general de mutantes <i>yap1/yap1b</i>	121
Figura Sup. 13 - Patrón de expresión en pez cebra de genes regulados por Yap1.....	122
Figura Sup. 14 - Longitud y anchura de embriones <i>yap1^{-/-} taz^{-/-}</i> de pez cebra	123
Figura Sup. 15 - Patrón de expresión de <i>ntl</i> en embriones <i>yap1^{+/+}</i> y <i>yap1^{-/-}</i> en embriones de medaka de estadio 19	124
Figura Sup. 16 - Inyección por separado de morfolinis contra <i>ctgfa</i> y <i>cyr61</i> en medaka	125
Figura Sup. 17 - Generación de alelos mutantes para el gen <i>ctgfa</i> de medaka	126
Figura Sup. 18 - Inyección asimétrica de zCtgfa en embriones de medaka para establecer gradientes.....	127
Figura Sup. 19 - Time lapse de la parte posterior de embriones de medaka durante la gastrulación	128
Figura Sup. 20 - Time lapse de la parte anterior de embriones de medaka durante la gastrulación	129
Figura Sup. 21 - Rastreo de trayectorias celulares en la región anterior de embriones de medaka durante la gastrulación.....	130

Tabla 1 - Picos identificados de iDamID-seq y genes asociados	71
Tabla 2 - Términos GO-Slim identificados de genes asociados a Yap1 en medaka	74
Tabla Sup. 1 - Lista de oligonucleótidos.....	131
Tabla Sup. 2 - Identificadores y genes de la familia Yap para diferentes especies.....	132
Tabla Sup. 3 - Lista completa de picos iDamID-seq de medaka y pez cebra	132
Tabla Sup. 4 - Términos de GO-Slim identificados de genes asociados a Yap1 y Yap1b en medaka.....	132
Tabla Sup. 5 - Datos de expresión de genes identificados en el RNA-seq	132
Tabla Sup. 6 - Genes reprimidos en mutantes <i>yap1</i> ^{-/-} de medaka asociados a picos iDamID-seq.....	133
Película Sup. 1 - Time-lapse de un embrión WT de medaka	134
Película Sup. 2 - Time-lapse de un embrión CRISPRante de <i>ctgfa</i> de medaka	134
Película Sup. 3 - Time-lapse de un embrión morfante de <i>ctgfa</i> y <i>cyr61</i> de medaka	134
Película Sup. 4 - Time-lapse detallado de la región posterior de un embrión WT de medaka	134
Película Sup. 5 - Time-lapse detallado de la región posterior de un embrión CRISPRante de <i>ctgfa</i> de medaka.....	134
Película Sup. 6 - Time-lapse detallado de la región posterior de un embrión morfante de <i>ctgfa</i> y <i>cyr61</i> de medaka	134
Película Sup. 7 - Time-lapse detallado de la región anterior de un embrión WT de medaka	135
Película Sup. 8 - Time-lapse detallado de la región anterior de un embrión CRISPRante de <i>ctgfa</i> de medaka.....	135
Película Sup. 9 - Time-lapse detallado de la región anterior de un embrión morfante de <i>ctgfa</i> y <i>cyr61</i> de medaka	135

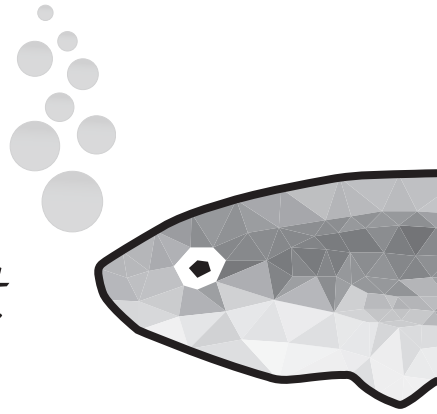
| *Abreviaturas*

BMP	Proteína morfogenética de hueso (B one M orphogenetic P rotein)
Cas9	C RISPR a ssociated protein 9
cDNA	Ácido desoxirribonucleico (ADN) codificante
CRISPR	C lustered R egularly I nterspaced S hort P alindromic R epeats
ctgfa	C onnective t issue g rowth f actor A
cyr61	C ysteine- r ich angiogenic inducer 61
DamID-seq	Identificación y secuenciación por ADN adenina metiltransferasa
ddPCR	<i>Digital Droplet PCR</i>
DEL	Capa profunda (D eep I ayer)
ECM	Matriz extracelular
EMT	Transición epitelio-mesénquima
EST	Marcador de secuencia expresada (E xpressed s equen ce t ag)
EVL	Capa envolvente (E nveloping I ayer)
FGF	Factor de crecimiento de fibroblastos (Fibroblast Growth Factor)
gDNA	Ácido desoxirribonucleico (ADN) genómico
GO	Ontología genética
H2B	Histona H2B
hpf	Horas post-fecundación
IPTG	Isopropil- β -D-tiogalactósido
ISH	Hibridación in situ
min	Minutos
mm	Milímetros
MMP	Metaloproteinasas de matriz
MO	Morfolino
mRNA	Ácido ribonucleico (ARN) mensajero
MSC	Células troncales mesenquimales
MSD	Desplazamiento cuadrático medio (M ean S quare D isplacement)

MZT	Transición materno-cigótica (M aternal-to- z igotic t ransition)
Ng	Nanogramos
NGS	Secuenciación de última generación (N ext G eneration S equencing)
nm	Nanometros
ntl	<u>no-tail</u>
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa
pH3	Fosfo-histona 3
px	Píxeles
qPCR	PCR cuantitativa
ROCK	Kinasas asociadas a Rho
SFK	Kinasas de la familia Src
SSC	Células troncales esqueléticas
Stg	Etapas (<i>Stage</i>)
Taz	<u>T</u>ranscriptional co-<u>a</u>ctivator with <u>PDZ</u>-binding domain
TEAD	<u>TEA</u> <u>D</u>omain family member
TF / FT	Factor de transcripción
TGF-β	Factor de crecimiento transformante β (T ransforming G rowth F actor β)
TSS	Sitio de inicio de la transcripción (T ranscription S tart S ite)
U. A.	Unidades Arbitrarias
WT	Estirpe silvestre (<u>W</u>ild <u>t</u>ype)
X-Gal	5-bromo-4-cloro-3-indol-β-D-galactopiranósido
Yap1	<u>Y</u>es-<u>a</u>ssociated <u>p</u>rotein-<u>1</u>
YSL	Capa sincitial del vitelo (<u>Y</u>olk <u>S</u>yncytial <u>L</u>ayer)
yg	Microgramos
μL	Microlitros
μm	Micras



Resumen / Abstract

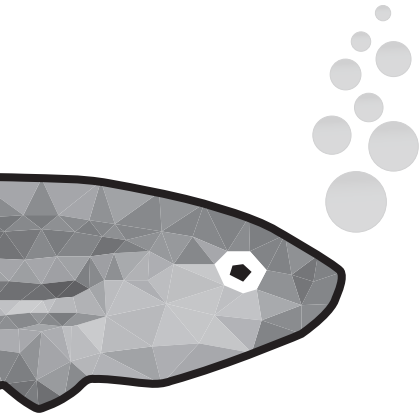


Yap y Taz son dos proteínas parálogas que pertenecen a una familia de activadores transcripcionales implicada en el correcto mantenimiento de tejidos adultos y en el proceso de embriogénesis, al coordinar procesos de proliferación, diferenciación celular y apoptosis mediante su unión en el núcleo a diversos factores de transcripción, fundamentalmente los pertenecientes a la familia Tead. Su actividad está controlada por tres mecanismos de regulación principales: la ruta Hippo, compuesta por una serie de kinasas que bloquean el ingreso de estos efectores al núcleo; la tensión morfogenética, que favorece el acceso de Yap y Taz al núcleo mediante un proceso estrictamente mecánico dependiente de la activación de integrinas y la polimerización de fibras de estrés; y la acción de kinasas de la familia Src/Yes (SFKs), que activan a estos parálogos mediante procesos de fosforilación dependientes e independientes de la ruta Hippo. El propósito principal de este trabajo consiste en describir la historia evolutiva de las proteínas de la familia Yap en vertebrados y analizar su papel morfogenético durante la gastrulación de teleósteos. El desarrollo de la presente tesis doctoral ha permitido demostrar la existencia de un tercer parólogo de la familia Yap, bautizado como *Yap1b*, que presenta una estructura divergente de Yap y Taz, con un dominio de activación transcripcional C-terminal que carece de residuos de tirosina y que por tanto escapa al control ejercido por kinasas SFK. El gen *yap1b*, que cuenta con una importante contribución materna pero cuyo nivel de expresión desciende súbitamente tras la activación del genoma cigótico, mantiene una relación epistásica negativa con el gen *yap1*, puesto que su mutación simultánea agrava el fenotipo de embriones mutantes *yap1*^{-/-} en medaka. La combinación de experimentos de iDamID-seq, RNA-seq, inmunofluorescencia, hibridación in situ e imaging demuestra que Yap desempeña un papel destacado en el establecimiento de los ejes principales del embrión de medaka y en el proceso de convergencia-extensión de células precursoras durante la gastrulación al promover la expresión de genes implicados en procesos de adhesión y migración celular, fundamentalmente genes que codifican para proteínas de la familia CCN como *ctgfa* y *cyr61*.

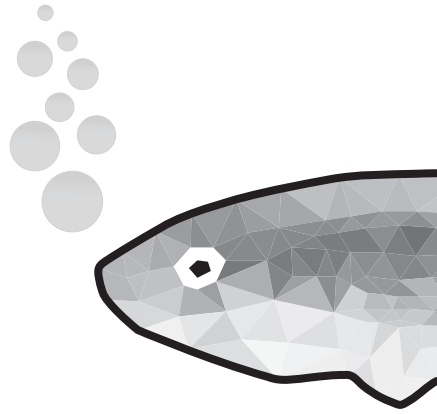
Palabras clave: CRISPR-Cas9, evolución, gastrulación, iDamID-seq, teleósteos, Yap, Yap1b.

Yap and Taz are two paralog proteins that belong to a family of transcriptional co-activators involved in tissue homeostasis and embryogenesis. They coordinate proliferation, cell differentiation and apoptotic processes through its binding, once translocated to the cell nucleus, to transcription factors, especially those from the Tead family. Essentially, their activity is controlled by three different upstream mechanisms: the Hippo signaling pathway, composed by a set of kinases that block the entrance of these effectors to the nucleus; morphogenetic tension, which promotes the access of Yap and Taz to the nucleus in a strictly mechanical process depends on integrin activation and stress fibers polymerization; and phosphorylation by Src/Yes kinases (SFKs), which activates these paralogs in both Hippo-dependent and independent manners. The main aim of this work consists of describing the evolutionary history of Yap family proteins in vertebrates, and analyzing their role as morphogenetic factors during the gastrulation in teleosts. The findings described in this Doctoral Thesis let us proof the existence of a third Yap paralog, here referred as *Yap1b*, which has a structure divergent from that of Yap and Taz at its C-terminal domain. Yap1b lacks the key tyrosine residues that can be phosphorylated by kinases of the Src family. The *yap1b* shows a strong maternal contribution, decreasing dramatically after zygotic genome activation. We suggest that *yap1* and *yap1b* show a negative epistatic interaction, as their simultaneous mutation aggravates the phenotypic defects described in *yap1*^{-/-} medaka mutant embryos. The combination of iDamID-seq, RNA-seq, immunofluorescence, in situ hybridizations and imaging experiments suggest that Yap plays a key role in establishing the main body axis of the embryo by controlling convergent-extension of progenitor cells during gastrulation. We show that Yap paralogs control a morphogenetic program, which includes the expression of genes involved in cell adhesion and migration, especially genes that encodes for proteins of the CCN family, such as *ctgfa* and *cyr61*.

Keywords: CRISPR-Cas9, evolution, gastrulation, iDamID-seq, teleosts, Yap, Yap1b.



Introducción



1. La familia Yap/Taz

1.1. Introducción

La familia Yap/Taz es un conjunto de coactivadores transcripcionales que desempeña un rol clave tanto durante la embriogénesis, en la que actúa como un regulador principal del crecimiento, la diferenciación y la supervivencia celular, como en el mantenimiento de tejidos y órganos adultos, interviniendo como actores principales en procesos de reparación tisular y en la progresión del cáncer (Piccolo et al., 2014; Varelas, 2014; Zanconato et al., 2016). *Yes-associated protein-1* (también conocido como *Yap1* o *Yap*) fue el primer miembro identificado de esta familia de cofactores, en el contexto molecular de su interacción física con las kinasas de tirosina de la familia Src/Yes (Sudol, 1994). Unos años más tarde, un segundo parólogo de esta familia de proteínas, *Taz* (*Transcriptional co-activator with PDZ-binding domain*, también conocido como *WW domain containing transcription regulator protein 1* o *Wwtr1*), fue identificado a través de su interacción con la proteína 14-3-3 (Kanai et al., 2000).

1.2. Estructura

Bajo su condición de proteínas parálogas, Yap y Taz comparten una estructura muy similar, establecida en una serie de dominios proteicos con una función muy definida: un dominio TEAD y dominios de unión a proteínas 14-3-3 amino-terminales, uno o dos dominios WW (caracterizados por la presencia de una pareja de aminoácidos de triptófano) y un dominio de activación transcripcional en su extremo carboxi-terminal rematado en un pequeño dominio PDZ de reconocimiento de secuencia (Figura 1). Yap y Taz, al igual que su ortólogo en *Drosophila* Yorkie (Huang et al., 2005), carecen de dominios de unión a DNA, por lo que precisan la incorporación de otros factores de transcripción para poder ejercer su función reguladora. Un estudio pionero reveló la interacción existente entre el dominio amino-terminal de Yap y los cuatro miembros de la familia de factores de transcripción Tead en vertebrados. Este estudio también determinó que, una vez formado el complejo Yap/Tead, el dominio carboxi-terminal de Yap se comporta como el auténtico dominio de activación transcripcional de Tead (Vassilev et al., 2001). Esta interacción fue confirmada posteriormente para los correspondientes ortólogos de Yap y Tead en *Drosophila*, Yorkie y Scalloped (Bandura y Edgar, 2008; Zhang et al., 2008). A pesar de que se ha estimado que la actividad de Yap/Taz puede ser controlada por multitud de factores de transcripción como AP-1, Runx, β -catenina, Tbx, p53 y Smad (Piccolo et al., 2014), se ha demostrado que sólo su interacción con Tead es

esencial para su unión indirecta a la cromatina y por tanto para su función transcripcional (Galli et al., 2015; Ze Li et al., 2010; Stein et al., 2015; Zanconato et al., 2015; Zhao et al., 2008). A excepción del factor Tead, que se une al dominio amino-terminal de Yap y Taz, otros cofactores, como p53 (Strano et al., 2001), Smad7 (Ferrigno et al., 2002) y Runx2 (Yagi et al., 1999) lo hacen a través de los dominios WW.

El dominio de unión a TEAD de Yap/Taz cuenta con varios motivos HXRXXS, que pueden ser fosforilados por kinasas reguladoras, fundamentalmente Lats. Asimismo, el dominio de transactivación cuenta con varios residuos de tirosina, fosforilables por kinasas Src/Yes. Se darán detalles de estas interacciones en los siguientes apartados.

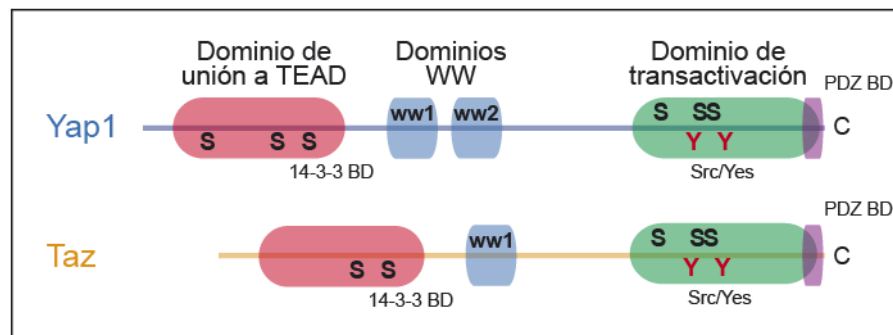


Figura 1 – Estructura básica de las proteínas Yap y Taz. Las proteínas de la familia Yap/Taz constan de cuatro dominios fundamentales: un dominio de unión a TEAD (rojo), uno o dos dominios WW (azul) y un dominio de activación transcripcional (verde), rematado en un motivo PDZ de reconocimiento de secuencia (violeta).

1.3. Regulación de la actividad de Yap y Taz

Una de las características más relevantes del complejo Yap/Taz-Tead es su capacidad para integrar la información transduccional procedente de rutas de señalización con aquella procedente de las propiedades estructurales de la célula, como la adhesión, la densidad, la polaridad, la contractilidad y la morfología celulares, y adecuarla al contexto de una respuesta transcripcional. Esta capacidad integradora se consigue a través de una serie de elementos de entrada (o *inputs*) de carácter regulador.

1.3.1. Ruta Hippo

De entre los diferentes elementos de entrada que regulan la actividad de las proteínas Yap y Taz, el más conocido es la ruta Hippo (Figura 2A). Los principales componentes de la ruta

Hippo fueron identificados gracias al llamativo sobrecrecimiento tisular observado en ciertos mutantes de *Drosophila* (Harvey et al., 2003; Tapon et al., 2002; Wu et al., 2003). Esta cascada señalizadora, conservada evolutivamente, está controlada fundamentalmente por dos kinasas que actúan consecutivamente, Mst1/2 y Lats1/2 (denominadas como Hippo y Warts en *Drosophila*, respectivamente), y que convergen en la fosforilación de Yap (en humanos, Lats fosforila a Yap en su residuo de serina S127) y de Taz, lo que resulta en su secuestro por parte de la proteína 14-3-3 y/o su posterior degradación en el citoplasma. Cuando esta ruta inhibidora no se encuentra activa, Yap y Taz pueden acceder libremente al núcleo y llevar a cabo su función como coactivadores transcripcionales (Meng et al., 2016).

Por otra parte, la función del núcleo principal de la ruta Hippo puede estar controlada por otras proteínas anexas. Uno de estos ejemplos es la proteína AMOT, que puede ser fosforilada por Lats, secuestrando a Yap e impidiendo tanto la polimerización de fibras de estrés como la formación de complejos de adhesión, lo que dificulta la migración de las células (Dai et al., 2013). La kinasa Lats también puede ser fosforilada y activada por la kinasa MAP4K (Zheng et al., 2015). Numerosas proteínas que forman parte de complejos de adhesión de la célula también parecen tener algún tipo de conexión con la ruta Hippo. Así, en *Drosophila*, las proteínas Mer, Ex y Kibra activan a la kinasa Wts (Lats) de manera cooperativa (Baumgartner et al., 2010).

1.3.2. Estrés mecánico

Además de la ruta Hippo, la actividad de Yap/Taz puede ser controlada por muchos otros factores, entre los que destaca el estrés mecánico (Figura 2B). En este sentido, la translocación de Yap/Taz al núcleo está bajo control de la geometría celular, de tal manera que permanece activa en células que tienden a extenderse por la superficie sobre la que se encuentran asentadas e inactiva en células redondas y compactas, en tejidos de alta densidad celular (Dupont et al., 2011). El grado de dureza de la superficie sobre la que se asienta la célula puede influir en su diferenciación. Así, las células troncales mesenquimales (MSC) pueden diferenciarse en adipocitos si el grado de dureza del medio sobre el que se asientan es bajo, o en osteoblastos si éste es alto (Engler et al., 2006; McBeath et al., 2004). La capacidad de Yap/Taz para captar la tensión morfogénica se basa en la contractilidad celular dependiente de miosina, en la polimerización de filamentos de actina y en el acoplamiento entre moléculas de integrina y talina. En otras palabras, la formación de fibras de estrés establece todo un entramado molecular mecanosensible (Aragona et al., 2013; Dupont et al., 2011; Elosegui-

Artola et al., 2016). Durante los últimos años, el estudio de la conexión entre la transmisión de fuerzas a través de la polimerización de fibras de estrés y la actividad de Yap/Taz ha generado una importante controversia, pues mientras que unos estudios han determinado que esta conexión es dependiente de la ruta Hippo (Wada et al., 2011), otros han adoptado la postura opuesta (Dupont et al., 2011). En este sentido, un estudio reciente ha demostrado la existencia de un acoplamiento mecánico entre la matriz extracelular y la envoltura nuclear que es responsable de la translocación directa de Yap/Taz a través de los poros nucleares en un proceso estrictamente mecánico que es, por tanto, totalmente independiente de la actividad de la ruta Hippo (Elosegui-Artola et al., 2017).

A pesar de la aparente independencia de este mecanismo de regulación con respecto a la ruta Hippo, existen ciertas proteínas que parecen conectar de algún modo la formación de fibras de estrés con algunos componentes de la ruta Hippo. En el apartado anterior se expuso el ejemplo de AMOT. La proteína Ajuba, identificada en *Drosophila*, secuestra a Lats en las uniones adherentes de la célula cuando ésta se encuentra sometida a una fuerte tensión morfogénica (Rauskolb et al., 2014).

1.3.3. Fosforilación por kinasas de la familia Src/Yes.

El segundo mecanismo independiente de la ruta Hippo que controla la actividad de Yap/Taz lo constituyen los miembros de la familia de kinasas de tirosina Src (SFKs) (Figura 2C). Como se mencionó líneas atrás, Yap se identificó originalmente como una proteína que actúa como socia de Yes y Src (Sudol, 1994). Se ha demostrado que cierta clase de SFKs, entre las que se encuentran las proteínas c-Abl, Src y Yes, fosforilan YAP en aminoácidos de tirosina específicos (especialmente la tirosina Y357 en humanos), localizados en su dominio de activación transcripcional carboxi-terminal (Levy et al., 2008; Rosenbluh et al., 2012; Vlahov et al., 2015; Zaidi et al., 2004). Más recientemente, se ha notificado la existencia de procesos de fosforilación en otros tres residuos conservados de tirosina (incluyendo la tirosina Y357) en keratinocitos (Li et al., 2016). Al contrario que el papel inhibitorio que juega la ruta Hippo, la fosforilación de Yap controlada por SFKs provoca la activación transcripcional de los complejos Yap/Tea (Calvo et al., 2013; Li et al., 2016; Tamm et al., 2011), Yap/Runx2 (Zaidi et al., 2004), Yap/p73 (Levy et al., 2008) y Yap/ β -catenina/Tbx5 (Rosenbluh et al., 2012; Vlahov et al., 2015). Como dato a recalcar, se ha sugerido que la fosforilación de Yap por mediación de Src podría constituir un modo de regulación de la actividad oncogénica de la célula ejercido por esta kinasa (Li et al., 2016), lo que se ajusta a la activación generalizada de Yap observada en células

cancerosas (Zanconato et al., 2016). La activación de kinasas Src, que se produce a través de la activación de las adhesiones focales, también puede regular negativamente la función de la ruta Hippo mediante fosforilación directa sobre Lats (Kim y Gumbiner, 2015).

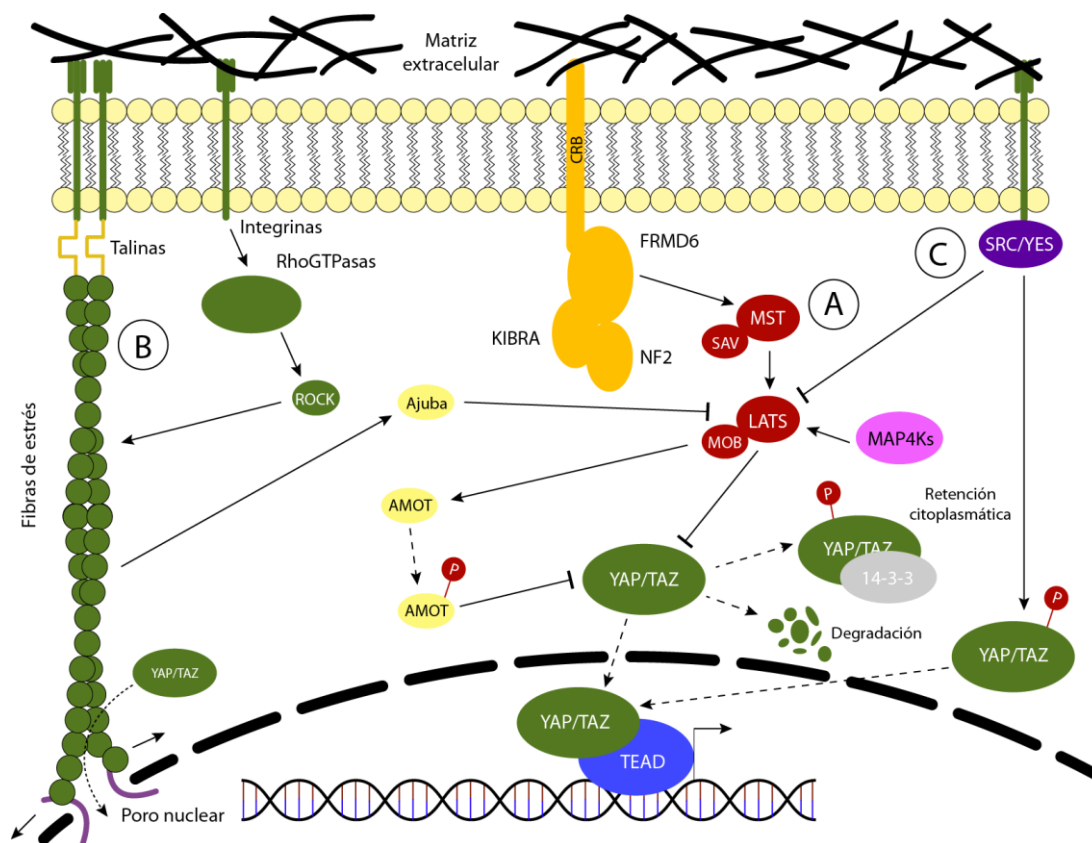


Figura 2 – Mecanismos de regulación de la actividad de Yap/Taz. A) **Ruta Hippo.** Las kinasas Lats fosforilan a Yap y Taz, provocando su confinamiento en el citoplasma y/o su posterior degradación, impidiendo su entrada al núcleo y su interacción con TEAD u otros cofactores de transcripción. Esta ruta puede ser a su vez controlada por otras vías y proteínas anexas. B) **Estrés mecánico.** La presión ejercida por el medio circundante promueve la formación de fibras de estrés a través de la activación de integrinas, que facilitan la entrada de Yap/Taz en el núcleo a través de un proceso estrictamente mecánico. C) **Kinasas Src/Yes.** Estas kinasas fosforilan a Yap y Taz en su dominio de transactivación y promueven su acceso directo al núcleo.

1.4. Yap y Taz a lo largo de la Evolución

Diversos análisis filogenéticos han permitido trazar la historia evolutiva del complejo Yap/Tead. Mientras que algunos ortólogos de Tead están ya presentes en levaduras, la existencia de proteínas de la familia Yap se remonta a los metazoos menos evolucionados, como los pertenecientes al grupo *Nematostella* (Hilman y Gat, 2011). Sin embargo, a pesar de esta conservación general, se ha descubierto la existencia de diferencias significativas entre especies. La conservación de residuos de tirosina requeridos para la fosforilación de Yap y

otros parálogos por parte de las SFKs parecen suponer una innovación en la región carboxi-terminal de las proteínas de la familia Yap/Taz en vertebrados (Elbediwy y Thompson, 2018; Hilman y Gat, 2011). Los parálogos *Yap* y *Taz* emergieron en vertebrados tras dos rondas consecutivas de duplicación del genoma completo (Dehal y Boore, 2005). Los datos recabados de diferentes modelos de vertebrados indican que existe un solapamiento parcial en la función ejercida por *Yap* y *Taz* durante el desarrollo. Mutaciones individuales de *Yap* y de *Taz* en ratones dan lugar a una amplia variedad de fenotipos: desde defectos en la elongación y en la vasculogénesis del saco vitelino, así como letalidad temprana en el estadio de desarrollo E8.5 en embriones mutantes *Yap*^{-/-}, hasta la aparición de ovarios poliquísticos y muerte prematura en mutantes adultos *Taz*^{-/-} (Morin-Kensicki et al., 2006; Xin et al., 2013). No obstante, el fuerte fenotipo observado en embriones mutantes dobles, que tienden a causar una elevada letalidad antes de la fase de mórula (Nishioka et al., 2009), y el resultado de *knockouts* dobles específicos de tejido (Deng et al., 2017; Lee et al., 2016; Singh et al., 2016) indican redundancia funcional para *Yap* y *Taz*. Se ha demostrado que dos parálogos de la familia Yap/Taz se encuentran presentes en el genoma de dos de los principales clados de teleósteos, representados por el pez cebra (*Danio rerio*, del grupo *Otocephala*) y el pez medaka (*Oryzias latipes*, perteneciente al grupo *Euteleosteomorpha*). Al igual que sucede en mamíferos, los dos parálogos de teleósteos de esta familia de proteínas parecen desempeñar funciones redundantes, provocando alteraciones graves en el desarrollo temprano cuando ambos se encuentran inactivos simultáneamente (Kimelman et al., 2017; Miesfeld et al., 2015).

2. Gastrulación en peces teleósteos

2.1. Introducción

La gastrulación es un proceso morfogénético embrionario de gran complejidad que ocurre en etapas tempranas del desarrollo de organismos multicelulares. La gran cantidad de mecanismos moleculares, celulares y tisulares que participan en esta etapa del desarrollo (revisado en Rohde y Heisenberg, 2007; Solnica-Krezel y Sepich, 2012) pone de manifiesto la complejidad de un proceso que resulta clave para la viabilidad del embrión, pues en él se producen dos hechos de importancia capital para ello: la diferenciación de las tres capas germinales (ectodermo, mesodermo y endodermo) y la formación de los ejes corporales principales del embrión (antero-posterior, dorso-ventral y lateral).

En los momentos previos al inicio de la gastrulación, el embrión de los peces teleósteos consiste en una masa de células que se han dividido de manera sincrónica, el *blastodermo*, ubicada en la superficie de la región superior del vitelo (*yolk*), en lo que se conoce como el *polo animal* del embrión. Está compuesto por dos tipos de blastómeros: las células que constituyen la *capa envolvente* o *enveloping layer* (EVL), un epitelio monoestratificado que recubre al resto de blastómeros, y las células de la denominada *capa profunda* o *deep layer* (DEL). En este punto ya se ha iniciado la activación de la transcripción cigótica, y los blastómeros comienzan a dividirse de manera asincrónica. Otra de las estructuras embrionarias que se forman durante esta fase es la *capa sincitial del vitelo* o *yolk syncytial layer* (YSL), ubicada en la región adyacente al blastodermo (Figura 4A). Esta capa surge como resultado de la fusión de los blastómeros más periféricos al vitelo (Kimmel y Law, 1985), constituyendo un sincitio que está involucrado en los primeros pasos de diferenciación que se producen durante la gastrulación (Koos y Ho, 1998).

2.2. Principales movimientos de gastrulación

A nivel tisular, la gastrulación es un proceso que engloba cuatro fases, que en ciertos momentos llegan a producirse simultáneamente: epibolia, internalización, convergencia y extensión (Figura 3 y Figura 4A-D).

La **epibolia** es un proceso morfogénético que implica la extensión isotrópica de un tejido al tiempo que éste reduce su grosor (Trinkaus y Lentz, 1967). En teleósteos, supone la expansión coordinada de todo el conjunto de células desde el polo animal hasta el polo opuesto (el *polo vegetal*) mediante movimientos de intercalación radial desde las capas más profundas a las más superficiales de la capa DEL. Mediante estos movimientos de intercalación, que carecen de direccionalidad definida (Bensch et al., 2013), las células más internas de la capa DEL pasan a ocupar las regiones más cercanas a las capas más superficiales, mezclándose con los blastómeros externos, aunque sin sobrepasar en ningún caso la capa EVL (Keller, 1980; Warga y Kimmel, 1990). Al tiempo que el tejido blastodérmico se expande por la superficie del embrión, las capas EVL e YSL hacen lo propio, de manera coordinada.

La **internalización** consiste en un proceso de invaginación, a través del blastoporo, de células precursoras del mesodermo y del endodermo por debajo de la capa de células que dará lugar al ectodermo cuando la epibolia ha alcanzado el 50% de la superficie del embrión. Una vez que se han internalizado, los precursores mesendodérmicos forman una capa interna

denominada *hipoblasto* que se desplaza hacia el polo animal al tiempo que la capa que no ha internalizado, el *epiblasto*, prosigue el proceso de epibolia hasta el polo vegetal. La separación entre estas dos capas se mantiene durante toda la gastrulación (Warga y Kimmel, 1990). Este proceso está regido por eventos de migración en el que las células se desplazan mediante una ingresión coordinada, pero como entidades individuales y no como un tejido verdaderamente cohesionado, al contrario que en otros mecanismos que tienen lugar durante la gastrulación (Solnica-Krezel, 2005; Kane y Adams, 2002). Las células que llevan a cabo el proceso de internalización están polarizadas inicialmente, por lo que, para poder internalizarse correctamente, deben someterse a un proceso celular de conversión conocido como *transición epitelio-mesénquima* (EMT) (Wu et al., 2007). Durante este proceso, las moléculas de adhesión celular y las uniones adherentes de las células epiteliales que darán lugar posteriormente al endodermo y al mesodermo se desensamblan, al tiempo que se produce una reorganización de su citoesqueleto (Thiery et al., 2009).

La **convergencia** y la **extensión** son procesos evolutivos altamente conservados que implican el adelgazamiento dorso-ventral de las incipientes capas germinales y su elongación a lo largo de todo el eje antero-posterior del embrión, respectivamente (Keller et al., 1985). El proceso de convergencia-extensión se inicia de manera simultánea a la internalización, cuando la epibolia ha alcanzado el 50% de la superficie del embrión, si bien es mucho más intenso cuando ésta alcanza aproximadamente el 70%, y está marcado por la compactación de células en la región dorsal del embrión, que genera el denominado *escudo embrionario*, el organizador de Spemann-Mangold de los teleósteos (Warga y Nüsslein-Volhard, 1998). En el organizador se expresan factores de transcripción que promueven la migración de células, el más conocido de los cuales es *gooseoid* (*gsc*) (Niehrs et al., 1993). En este punto, la población de progenitores mesodérmicos se divide en tres: las células mesodérmicas axiales, que formarán estructuras tales como la placa precordial y la notocorda; las células mesodérmicas paraxiales, que constituirán los futuros somitas; y las células mesodérmicas laterales, que darán lugar a la placa lateral (Kimmel et al., 1990).

A nivel celular, la clave del doble proceso de convergencia y extensión reside en una combinación de movimientos de intercalación planar medio-lateral que permiten a las células replegarse hacia la futura línea media del embrión, y en un conjunto de divisiones celulares polarizadas para promover la extensión del tejido (Sepich et al., 2005). Sin embargo, el proceso de convergencia-extensión no es completamente homogéneo. Los progenitores de la placa precordial migran directamente al polo animal mediante la generación de protrusiones que les permiten desplazarse de manera autónoma (Ulrich et al., 2005), usando al epiblasto como

soporte sobre el que desplazarse (Montero, 2005). Las células del mesodermo lateral también migran de manera directa al polo animal al inicio del proceso de convergencia-extensión. Sin embargo, cuando la epibolia ha alcanzado el 70% de la superficie del embrión, estas células comienzan a exhibir un proceso de intercalación mediolateral más intenso, que las redirige a la región dorsal del embrión (Sepich y Solnica-Krezel, 2005). También se han identificado procesos de migración celular aleatoria que también contribuyen al fenómeno de extensión (Pézeron et al., 2008). Por último, en las regiones más ventrales, las células mesodérmicas tienden a migrar hacia el polo vegetal (Myers et al., 2002).

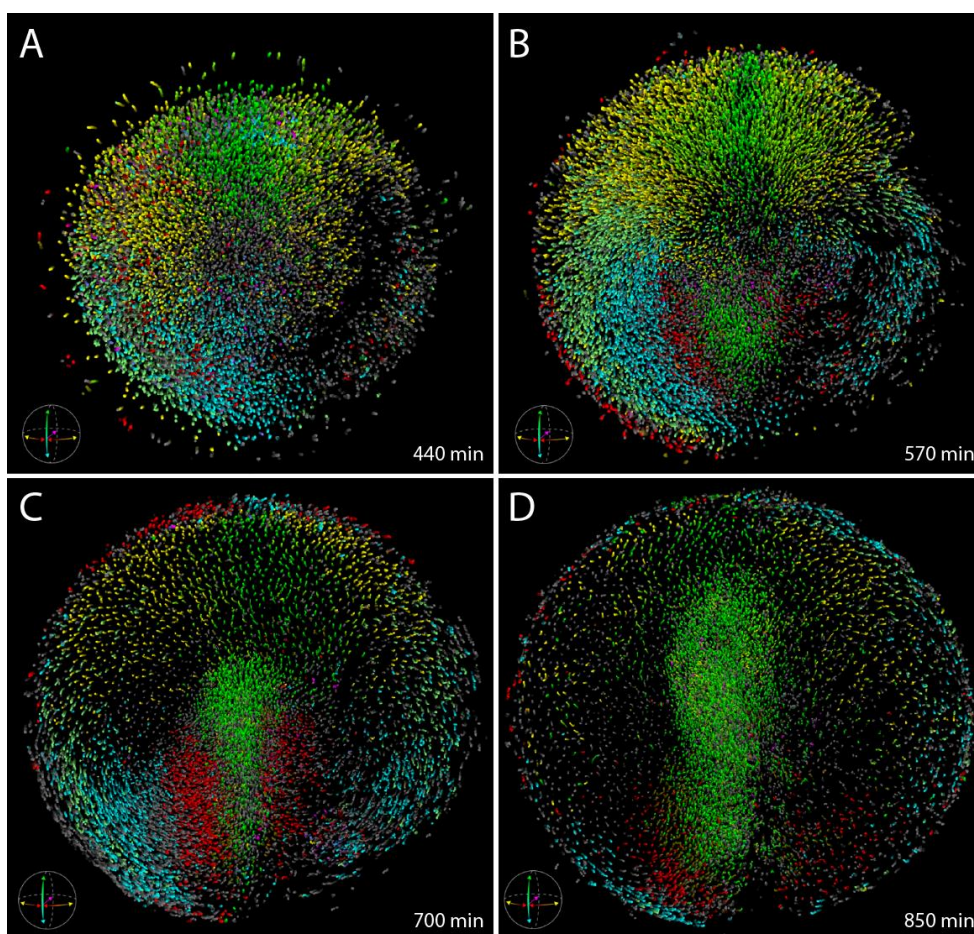


Figura 3 – Principales movimientos celulares registrados durante la gastrulación del pez cebra. **A)** Para cuando el embrión ha activado su transcripción cigótica, éste está compuesto de una masa de células que se acumulan en su polo animal. En este punto se inicia la epibolia: la capa de blastómeros más profunda se aproxima a la región más superficial mediante procesos de intercalación radial (puntos verdes y amarillos), al tiempo que el resto de células se desplazan al polo vegetal (puntos azules). **B)** Cuando el frente de epibolia ha alcanzado el 50% de la superficie del embrión, los precursores mesendodérmicos se invaginan (puntos rosas) e inician el proceso de convergencia (puntos rojos) y extensión (puntos verdes de la mitad inferior de la imagen). **C)** A medida que la epibolia progresa, el proceso de convergencia-extensión se va haciendo más evidente, especialmente en la futura región posterior del embrión. **D)** Cuando la epibolia ha finalizado, disminuyen los procesos de convergencia y el eje principal del embrión ya resulta reconocible. Imágenes tomadas del material suplementario de Keller et al. (2008).

2.3. Rutas de señalización implicadas en la gastrulación

Cuando una célula migra, ocurre un proceso dinámico que consta de cuatro fases. Primero, las células se polarizan debido a la captación de estímulos externos. Luego, esta polarización genera cambios en la organización del citoesqueleto, que genera protrusiones. Posteriormente, dichas protrusiones se anclan a soportes con una alta adherencia entre células o entre célula y matriz extracelular. Por último, un movimiento de retracción provoca el desplazamiento de la célula (Ridley et al., 2003). De forma similar, los movimientos de intercalación entre células vecinas y los cambios de morfología celular también comportan la reorganización del citoesqueleto y las uniones adherentes. Así pues, una correcta gastrulación requiere una coordinación altamente precisa de procesos de adhesión y migración/desplazamiento celular, que está determinada por la activación de diversas rutas bioquímicas y cascadas de señalización intracelular. Este férreo control ejercido por parte de la combinación de varios mecanismos intracelulares (Figura 4E) pone de manifiesto la complejidad de la gastrulación en teleósteos.

2.3.1. Ruta WNT

Wnt es una glicoproteína extracelular que se une a diversos receptores de membranas, entre los que destacan los de la familia Frizzled (Fz). Cuando esto sucede, activa una cascada de señalización intracelular que regula polaridad celular y migración. Numerosos estudios demuestran que la vía Wnt constituye el principal regulador del proceso de convergencia-extensión durante la gastrulación de los vertebrados, ya que muchos de sus componentes se han relacionado con esta etapa del desarrollo (Rauch et al., 1997; Heisenberg et al., 2000; Sepich et al., 2000; Topczewski et al., 2001).

La señal Wnt puede transducir a través de tres rutas principales: la ruta Wnt canónica, la ruta Wnt/PCP no canónica y la ruta Wnt/Ca²⁺.

En la **ruta Wnt canónica**, la unión de Wnt a sus receptores provoca la activación de la proteína citoplasmática Dvl, que inhibe la función del complejo APC-Axin-GSK3-CK1. Cuando este complejo está activo, provoca la degradación de la proteína β -catenina. De este modo, la activación de la ruta disocia el complejo, propiciando su acumulación citoplasmática y su posterior translocación al núcleo, donde forma complejos con otras proteínas que promueven la activación de la expresión de genes diana de la ruta (revisado en Moon et al., 2004). La translocación al núcleo del co-regulador transcripcional β -catenina promueve la expresión de

factores implicados en la formación del organizador embrionario (revisado en Langdon y Mullins, 2011).

La **ruta Wnt/PCP** constituye uno de los responsables principales de la adquisición de polaridad de la célula. Así, en *Drosophila*, la correcta distribución asimétrica de los componentes principales de esta vía de señalización determina la polarización de las células epiteliales (Strutt y Strutt, 2005). En el contexto de la gastrulación, la ruta Wnt/PCP se activa mediante la unión de Wnt5 y Wnt11 a los receptores Fz. Esto produce la formación de un complejo Dvl-Daam-RhoA (Habas et al., 2003), que induce la activación de kinasas asociadas a Rho (ROCK), lo que produce una reorganización del citoesqueleto que favorece la elongación y orientación mediolateral de la célula así como la formación de protrusiones, promoviendo movimientos de convergencia-extensión (Marlow et al., 2002; Wallingford et al., 2000). También se han identificado otras proteínas, como Prickle y Strabismus, que también desempeñan una función de importancia en estos movimientos de convergencia-extensión (Takeuchi et al., 2003; Carreira-Barbosa et al., 2003).

La **ruta Wnt/Ca²⁺** también se activa durante el desarrollo embrionario (Kuhl et al., 2000). La interacción entre Wnt y Fz libera grandes cantidades de Ca²⁺ intracelular, que activa a las kinasas dependientes de calcio (CamKII) y a la kinasa PKC, ambas implicadas en migración y adhesión celular durante la gastrulación (Kuhl et al., 2000; Sheldahl et al., 2003; Slusarski et al., 1997; Westfall et al., 2003). La proteína Cdc42 también podría considerarse como un efector de la ruta puesto que el bloqueo de su actividad en *Xenopus* provoca defectos similares a los observados por una sobreactivación de la misma (Choi y Han, 2002). La ruta Wnt/Ca²⁺ interacciona funcionalmente con la ruta Wnt/PCP durante la gastrulación (Kinoshita et al., 2003; Sheldahl et al., 2003).

2.3.2. Ruta Nodal

La proteína Nodal pertenece a la superfamilia de proteínas TGF- β , que transduce señales a través de receptores de serina-treonina kinasa (Whitman, 2001), promoviendo la formación de complejos Smad2-P/Smad4 que se translocan al núcleo y activan a factores de transcripción como FoxH1 y p53 (Schier, 2003). Estos factores de transcripción regulan la expresión de genes importantes en el proceso de convergencia-extensión. Entre ellos, destaca la molécula de miRNA *mir206*, que favorece la expresión de E-cadherinas, necesarias para que la gastrulación

se produzca correctamente (Liu et al., 2013). La activación de esta ruta es clave para la especificación del mesodermo y del endodermo (Feldman et al., 1998; Sampath et al., 1998).

2.3.3. Ruta BMP

Las proteínas morfogenéticas de hueso (*Bone Morphogenetic Proteins*, BMP) son necesarias para la correcta diferenciación de células ventro-laterales durante la especificación del eje dorso-ventral y por tanto, también desempeñan una función importante en la convergencia-extensión que tiene lugar durante la gastrulación. Como Nodal, las proteínas BMP también forman parte de la superfamilia de proteínas TGF- β . Durante la gastrulación del pez cebra, tres genes *bmp* (*bmp2b*, *bmp4* y *bmp7*) se expresan formando un gradiente dorso-ventral, con un mayor nivel de expresión en las regiones más ventrales. Este gradiente de BMP provoca que las células que captan esta señal tengan un nivel de adhesión célula-célula diferente en función de su posición, siendo más bajo cuanto mayor sea el nivel de BMP. El nivel de adhesión celular influye en el grado de migración, dependiente de la formación de lamelipodios, de las células del mesodermo lateral. Así, en las regiones ventro-laterales, donde la concentración de BMP es más baja, las células tienden a migrar hacia la región dorsal, mientras que las células ubicadas en la posición más ventral del embrión, donde la concentración de BMP es más alta, no experimentan fenómenos de convergencia-extensión (von der Hardt et al., 2007). La expresión del factor de crecimiento de fibroblastos (*Fibroblast Growth Factor*, FGF) en la región dorsal del embrión contribuye al establecimiento del gradiente dorso-ventral de BMP (Fürthauer, 2004).

2.3.4. Otras rutas de señalización

Existen otras muchas rutas y moléculas de señalización que también participan en la gastrulación de peces teleósteos. La kinasa **PI3K** puede controlar el movimiento celular activando de manera local pequeñas GTPasas y proteínas con dominios PH que regulan la dinámica del citoesqueleto, contribuyendo a la formación de protrusiones polarizadas (Merlot y Firtel, 2003). La actividad de la kinasa PI3K está controlada por el factor de crecimiento **PDGF**, una molécula señalizadora que controla procesos de proliferación, migración y formación de tejidos (Hoch y Soriano, 2003). Este factor favorece la migración de los progenitores mesodérmicos axiales de la placa precordial hacia el polo animal (Montero et al., 2003). La ruta **Jak/Stat** también controla el movimiento de las células durante la gastrulación.

Por ejemplo, la sobreexpresión de formas dominantes negativas de la kinasa Jak1 en pez cebra se traduce en un retraso en la epibolia y en embriones con un eje antero-posterior más corto y más ancho, indicios claros de defectos en la gastrulación (Conway et al., 1997). Por otra parte, la expresión de Stat3 es necesaria para una correcta convergencia-extensión y para que los precursores de la placa precordial puedan migrar adecuadamente al polo animal (Yamashita et al., 2002). La ruta **Eph-Ephrinas**, que regula procesos de adhesión celular y de reorganización del citoesqueleto para controlar el grado de atracción o repulsión entre células o entre éstas y la ECM (Murai y Pasquale, 2005), la ruta **Slit-Robo**, que controla procesos de intercalación medio-lateral del mesodermo axial típicos de la fase de convergencia-extensión (Yeo et al., 2001), y la ruta **Notch**, que participa en la especificación del endodermo (Kikuchi et al., 2004) son otros ejemplos de rutas de señalización implicadas en gastrulación.

2.4. Mecanismos de adhesión implicados en la gastrulación

Los cambios morfológicos que ocurren en la célula, así como los procesos de migración e intercalación que tienen lugar durante la gastrulación dependen de la regulación de la adhesión celular y de la actividad de su citoesqueleto (Figura 4E).

La adhesión intercelular, principalmente controlada por la formación de complejos adherentes de cadherinas, constituye un mecanismo tisular que está involucrado en la separación de las capas germinales, en la intercalación radial de las células durante la etapa de epibolia, así como en los procesos de transición epitelio-mesénquima y en la migración dorsal del mesodermo durante la gastrulación. La mutación del gen *cdh1*, que codifica para E-cadherinas, provoca retrasos en la epibolia y defectos en la convergencia-extensión en embriones de pez cebra al afectar a los procesos de intercalación radial y medio-lateral, dando como resultado embriones con un eje antero-posterior más corto y somitas más anchos (Shimizu et al., 2005). Las moléculas de α E-catenina, que en combinación con moléculas de β -catenina enlazan el citoesqueleto de actina celular a las uniones adherentes de la célula (revisado en Benjamin y Nelson, 2008), también juegan un papel importante en la adhesión célula-célula y en la migración celular, y en combinación con las moléculas de E-cadherina, también controlan la correcta intercalación radial de los blastómeros de la capa DEL (Schepis et al., 2012). Por otra parte, en los embriones de peces cebra, el grado de tensión superficial que capta la célula aumenta en las capas más superficiales y disminuye progresivamente a medida que se profundiza en el hipoblasto (Krieg et al., 2008). Esta diferencia en el grado de tensión entre precursores constituye uno de los factores que explica no sólo la correcta

separación de capas germinales, sino también el origen molecular de otros procesos como la delaminación, migración y elongación de las células precursoras mesendodérmicas, ya que la tensión generada por el citoesqueleto de actina favorece el aumento de la cohesión del tejido al fomentar una mayor tasa de expresión de E-cadherinas (Montero, 2005; Schötz et al., 2008).

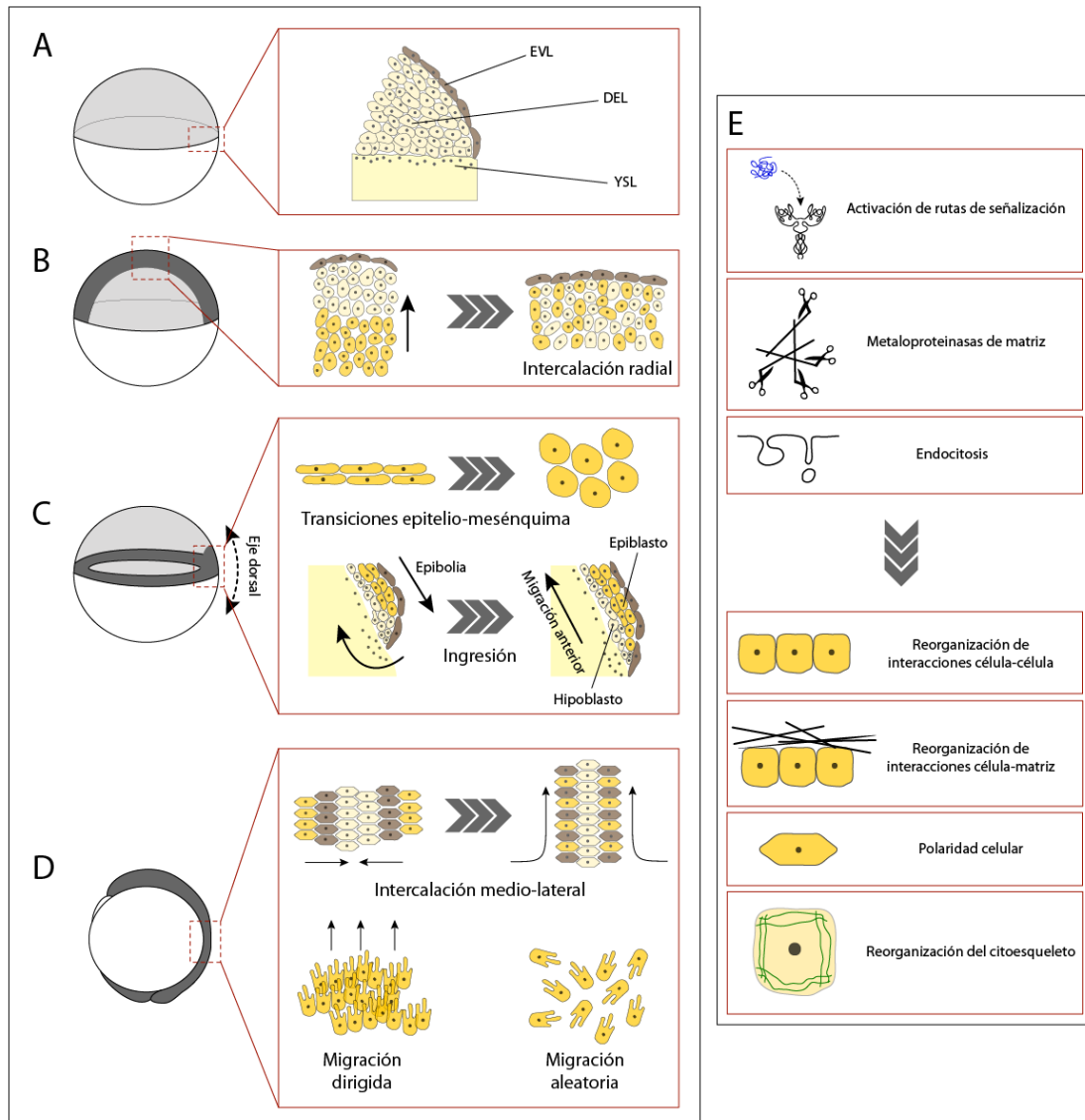


Figura 4 – Resumen de los principales mecanismos celulares implicados en la gastrulación de los teleosteos. **A)** Al comienzo de la gastrulación, el embrión es una masa de blastómeros (DEL) rodeada de un epitelio (EVL) y asentada sobre un sincitio (YSL) que la mantiene separada del vitelo. **B)** Durante la epibolia, los blastómeros más profundos de la capa DVL se desplazan hacia las capas superiores mediante procesos de intercalación radial. **C)** Cuando la epibolia ha alcanzado el 50% de la superficie del embrión, los precursores mesendodérmicos son sometidos a transiciones epitelio-mesénquima. Esto promueve su invaginación por debajo de la capa de células que prosigue la epibolia, generándose en el proceso dos capas diferenciadas: el epiblasto y el hipoblasto. Las células del hipoblasto comienzan a migrar hacia el polo animal. **D)** Al tiempo que se produce la invaginación de células mesendodérmicas, éstas comienzan a converger hacia la región media y dorsal del futuro embrión combinando procesos de intercalación medio-lateral, migración dirigida y migración aleatoria. **E)** El comportamiento que se aprecia en cada tejido durante la gastrulación se debe a fenómenos que ocurren a nivel celular (reorganización de adhesiones y del citoesqueleto, polarización celular...) que a su vez están controlados por una amplia diversidad de rutas bioquímicas y de señalización intracelular. Imagen modificada de Rohde et al. (2007).

La matriz extracelular (ECM) constituye un entramado de glicoproteínas que proporciona un soporte apto para la migración celular y la transmisión de fuerzas, por lo que juega un papel muy importante en la gastrulación de los vertebrados. El establecimiento de interacciones célula-ECM a medida que la gastrulación progresa resulta imprescindible para el éxito de este proceso. Uno de los principales componentes de la ECM es la fibronectina, que se distribuye formando una red de fibras que interacciona con los receptores de integrinas (especialmente los del tipo $\alpha 5 \beta 1$) de las células adyacentes, provocando a su vez la subsiguiente contracción de su citoesqueleto de actina, que permanece conectado a dichos receptores (revisado en Mao y Schwarzbauer, 2005). La combinación de fibronectina y este subtipo específico de integrinas resulta imprescindible para la correcta migración de los precursores mesodérmicos durante la fase de convergencia-extensión (Davidson et al., 2006). Asimismo, el bajo grado de adhesión célula-ECM favorece la migración aleatoria de los precursores del endodermo, mientras que los precursores ectodérmicos, con un grado de adhesión superior, se desplazan como un tejido epitelial, más cohesionado (Nair y Schilling, 2008). Al igual que la presencia de uniones adherentes intercelulares, la precisa arquitectura de la ECM contribuye a la correcta separación entre capas germinales. El epiblasto y el hipoblasto están separados firmemente por una capa de matriz extracelular rica en fibronectina, y el mesendodermo se mantiene separado del vitelo mediante una capa extra de matriz extracelular compuesta principalmente de fibronectina y laminina (Latimer y Jessen, 2010).

La matriz extracelular no es una estructura fija, sino que es remodelada periódicamente por metaloproteinasas de matriz (MMPs). Estas metaloproteinasas son endopeptidasas que degradan o modifican algunos componentes de la matriz extracelular, como las fibronectinas, las lamininas y diversos tipos de colágeno (revisado en Lynch y Matrisian, 2002). Una de las más importantes es la MMP de membrana de tipo 1 (MMP14), cuya función se ha demostrado como necesaria para controlar la movilidad celular durante la gastrulación. La metaloproteinasa MMP14 controla la correcta polarización de las células mesodérmicas, necesaria para su migración durante la etapa de convergencia-extensión. El grado de actividad de dicha metaloproteinasa está relacionado con la actividad de la ruta Wnt/PCP, ya que uno de sus componentes, la proteína transmembrana Vangl2, controla mediante endocitosis los niveles de la MMP14 en la superficie celular (Coyle et al., 2008; Williams et al., 2012). No es esta la única relación existente entre componentes de adhesión y de la ruta Wnt. En pez cebra, la pérdida de la proteína Prickle1a supone un descenso en los niveles de fibronectina del embrión durante la gastrulación, mientras que la mutación del gen *glypican4*, que codifica

para un receptor de WNT, aumenta el nivel de expresión de cadherinas, y por tanto, el nivel de adhesión célula-ECM (Dohn et al., 2013).

Para que las células puedan desplazarse sobre un sustrato cohesionado, las células deben ensamblar y desensamblar dinámicamente sus uniones adherentes, sean complejos de integrinas o de cadherinas. Procesos de endocitosis contribuyen a ello. Así, para que se produzcan correctamente los movimientos de intercalación-radial de la epibolia y las transiciones epitelio-mesénquima que tienen lugar durante la gastrulación, es necesario que se mantenga un estricto control dinámico del tráfico de E-cadherinas por toda la membrana, que se consigue mediante la endocitosis de estas proteínas (Bryant y Stow, 2004; Song et al., 2013). Este proceso de endocitosis también está conectado con las rutas de señalización y polarización celular. Un estudio previo revela que la ruta Wnt/PCP, por medio de la unión de Wnt11 a sus receptores, controla el tráfico de E-cadherinas mediante endocitosis, regulando el grado de cohesión del tejido mesendodérmico (Ulrich et al., 2005).

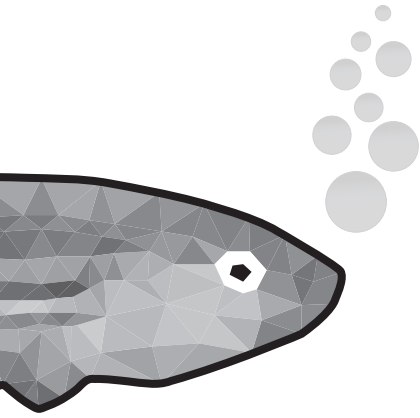
3. Conexiones entre la actividad de la familia Yap/Taz y mecanismos implicados en la gastrulación de teleósteos

Como se ha mencionado previamente, la vía de señalización que controla la actividad de las proteínas de la familia Yap/Taz es bastante intrincada. Este nivel de complejidad aumenta todavía más si la situamos en el contexto de la gastrulación de teleósteos, porque existen estudios que establecen conexiones entre estas proteínas efectoras y componentes celulares o vías de señalización que participan en esta etapa crítica del desarrollo.

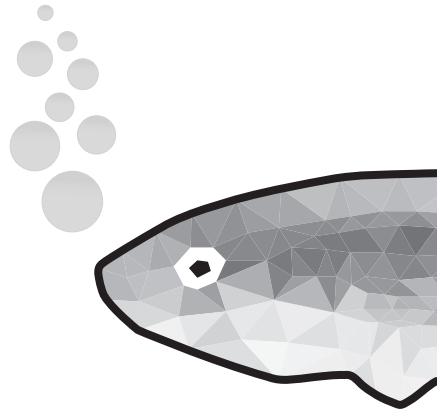
El ejemplo más claro lo encontramos en la ruta Wnt. Según experimentos recientes, Yap y Taz forman parte del complejo APC-Axin-GSK3-CK1 que induce la degradación de las moléculas de β -catenina (Azzolin et al., 2014). Por esta razón, cuando la ruta Hippo está activa, Lats2 fosforila a ambos cofactores y éstos quedan confinados en el citoplasma, donde nada les impide pasar a formar parte de dicho complejo. La acumulación de Taz en el citoplasma también impide la fosforilación y posterior activación de Dvl, lo que libera al complejo APC-Axin-GSK3-CK1 para que pueda ejercer su función represora (Varelas et al., 2010). Lats2 también puede bloquear la interacción de las moléculas de β -catenina con su cofactor Bcl9 (J. Li et al., 2013) y se cree que la proteína Amotl2a, que interacciona físicamente con Yap (Agarwala et al., 2015), puede inhibir la vía de señalización Wnt/ β -catenina durante el

desarrollo temprano de los peces cebra impidiendo que la β -catenina se transporte al núcleo (Li et al., 2012).

La modificación del grado de adhesión célula-ECM también afecta a la actividad de Yap y Taz. La actividad de la metaloproteinasa MMP14 descrita en el apartado anterior provoca cambios conformacionales que activan a la subunidad $\beta 1$ de las integrinas de células troncales de hueso (SSC) (Tang et al., 2013). Esto produce la reorganización del citoesqueleto de la célula y el aumento de su rigidez pericelular por medio de la activación de RhoGTPasas y kinasas ROCK (Eyckmans et al., 2011). La polimerización de fibras de estrés resultante favorece la entrada de los efectores Yap y Taz en el núcleo (Tang et al., 2013). Por tanto, la actividad de estas metaloproteinasas influye de manera directa en la actividad de Yap y Taz, si bien no ha podido demostrarse esta conexión en el contexto de la gastrulación de teleósteos. La activación de Yap y Taz también favorece la formación de adhesiones focales y la estabilidad del citoesqueleto por medio de la expresión de genes que codifican para proteínas que forman parte de estos elementos celulares. Ambos parálogos son activados debido a la presencia de fibras de estrés, aunque un mayor número de adhesiones focales no implica necesariamente una mayor tasa de translocación al núcleo (Nardone et al., 2017).

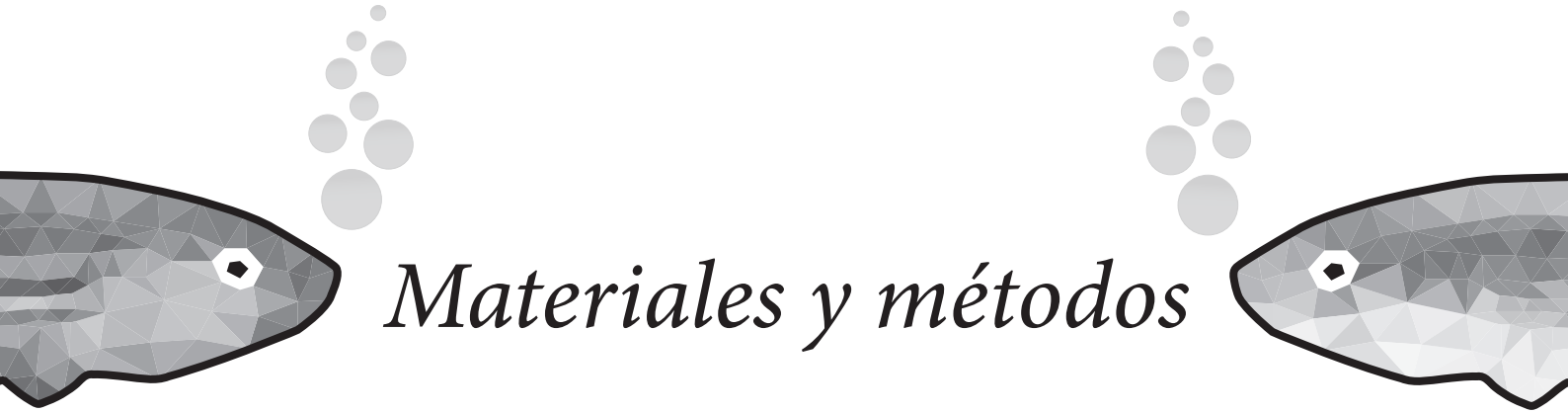


Objetivos



Los objetivos propuestos para el desarrollo de la presente tesis doctoral son los siguientes:

1. Investigar la historia evolutiva de los parálogos de la familia Yap en teleósteos.
2. Identificar las dianas de unión al genoma de los diferentes parálogos de la familia Yap durante la gastrulación de peces cebra y medaka.
3. Determinar el papel morfogénético llevado a cabo por los parálogos de la familia Yap en la gastrulación en teleósteos.
4. Analizar en detalle la función desempeñada por genes controlados por Yap1 durante la gastrulación de peces teleósteos.



Materiales y métodos

1. Mantenimiento de líneas de peces teleósteos

Como modelo experimental para el desarrollo de este proyecto, se optó por el uso de embriones de pez cebra (*Danio rerio*) y medaka (*Oryzias latipes*), dos especies de peces teleósteos separados por una distancia evolutiva de casi 150 millones de años. Ambos modelos destacan por la elevada cantidad de embriones fecundados que liberan por puesta, así como por la alta transparencia y rápido desarrollo de sus embriones (del orden de tres días para el pez cebra y del orden de ocho para el pez medaka), que posibilitan el planteamiento de una infinidad de estudios y análisis de morfogénesis y organogénesis durante el desarrollo temprano.

La estirpe silvestre de pez cebra *AB/Tübingen* (AB/TU) utilizada para el desarrollo de los análisis llevados a cabo en este trabajo se mantuvo bajo los estándares recogidos en los procedimientos descritos por Kimmel et al. (1995) y por la Zebrafish Model Organism Database (<http://zfin.org>; Sprague et al., 2003). Todas las etapas de desarrollo especificadas en este estudio (Figura Suplementaria 1) se muestran en horas post-fecundación (hpf) aplicadas a embriones incubados a 28.5°C según las pautas establecidas (Kimmel et al., 1995).

Asimismo, la estirpe silvestre de pez medaka *ICab* empleada a lo largo de este estudio se mantuvo bajo las condiciones experimentales propuestas por Iwamatsu (2004). Las etapas de desarrollo embrionario especificadas en cada uno de los experimentos descritos (Figura Suplementaria 2) se clasifican atendiendo a sus rasgos morfológicos (Iwamatsu, 2004).

Todos los embriones se mantuvieron en crecimiento en medio E3, que cuenta con las condiciones de sal y pH recomendadas para el mantenimiento de peces cebra y medaka.

2. Análisis filogenético de genes de la familia Yap

Las diferentes proteínas de la familia Yap se alinearon empleando la herramienta MAFFT (Boc et al., 2012). Los árboles de máxima verosimilitud se trazaron empleando el software PHYML empleando el modelo WAG +G y utilizando 100 iteraciones. Para obtener las secuencias proteicas y reconstruir el origen genómico de los diferentes miembros de la familia Yap, se utilizaron las anotaciones genómicas disponibles en las bases de datos NCBI (<https://www.ncbi.nlm.nih.gov/genome/>), UCSC (<http://genome.ucsc.edu/>), y Ensembl (<http://www.ensembl.org/index.html>). Se emplearon las siguientes versiones genómicas: *Anguilla japonica* (Anguila japonesa) *Ajaponica_ver_D2*; *Astyanax mexicanus* (Tetra ciego

mexicano) 2.0; *Clupea harengus* (Arenque común) ASM96633v1; *Danio rerio* (Pez cebra) Zv9; *Esox lucius* (Lucio europeo) EsoLuc1.0; *Gadus morhua* (Bacalao común) gadMor1; *Gasterosteus aculeatus* (Pez espinoso) BROAD_S1; *Homo sapiens* (Ser humano) hg38; *Lepisosteus oculatus* (Catán pinto) LepOcu1; *Mus musculus* (Ratón) m38; *Oreochromis niloticus* (Pez tilapia) Orenil1.0; *Oryzias latipes* (Pez medaka) v1.0; and *Paramormyrops kingsleyae* (Pez elefante) Pkings_v0.1.

3. Generación de líneas mutantes en pez medaka

3.1. Generación de línea mutante *yap1* en medaka. Se utilizó la tecnología CRISPR-Cas9 para mutar el exón 2 del gen *yap1* (Ensembl ID: ENSORLG00000002708). Concretamente, se utilizó la herramienta CCTop (Stemmer et al. 2015) para la identificación de regiones susceptibles de ser reconocidas por tres moléculas de sgRNA, que se generaron utilizando sendas parejas de oligonucleótidos (oligonucleótidos 1-2, 3-4 y 5-6 de la Tabla Suplementaria 1). Cada pareja de oligonucleótidos se hibridó entre sí para generar dímeros que fueron insertados mediante ligación en plásmidos pDR274 (*Addgene*, #42250), previamente linealizados con la enzima de restricción BsaI. Para llevar a cabo este proceso, se utilizaron aproximadamente 10 ng de cada dímero y 100 ng de vector, que se incubaron en presencia de ligasa T4 en un volumen final de 10 µL durante toda la noche a 16°C. Se transformaron 2 µL de este producto de ligación en bacterias *E. coli* DH5α competentes, que fueron sembradas en placas de kanamicina. Para confirmar el éxito de la ligación, se llevaron a cabo varias PCR de colonias utilizando el oligonucleótido universal M13Fw (que hibrida en el plásmido pDR274, en posición 5' junto al lugar de inserción del dímero) y el oligonucleótido reverso de cada dímero. Las colonias que resultaron positivas fueron procesadas con un kit de extracción de plásmidos (Miniprep Kit, *Sigma*) y fueron enviadas a secuenciar. Tras la confirmación de los resultados de secuenciación, los plásmidos obtenidos se linealizaron con la enzima HindIII y se purificaron por columna (GFX PCR DNA and Gel Band purification Kit, *G&E*).

Mediante transcripción in vitro (T7 MAXIscript Kit, *Ambion*) se sintetizó mRNA sin caperuza ni cola de poli(A) de cada sgRNA. Para la síntesis de mRNA de la proteína Cas9, el plásmido pCS2-nCas9n (*Addgene*, #47929) fue linealizado con NotI y purificado por columna. Se procedió a llevar a cabo una transcripción in vitro para generar mRNA de esta proteína partiendo desde el promotor SP6 presente en el plásmido (SP6 mMessage mMachine Kit, *Ambion*). En ambos casos, se emplearon de partida 1-1.5 µg de cada plásmido digerido. Tras

precipitar todos los productos de transcripción en presencia de EtOH y LiCl, éstos fueron limpiados utilizando el kit RNA cleanup Kit (*Qiagen*) y resuspendidos en 12 µL de agua libre de nucleasas.

Se inyectaron aproximadamente 5 nL de una mezcla de proteína de Cas9 a una concentración final de 300 ng/µL y de las tres moléculas de sgRNA diseñadas a una concentración final de 40 ng/µL en embriones de medaka en estadio de una célula utilizando un microinyector manual (*Narishige*) y agujas de vidrio (Glass Capillary Filament 1 mm x 58 mm x 6'', *CIBERTEC601500*).

Se procedió a la identificación de ejemplares *founder* (que presentan mutaciones en su línea germinal y que por tanto la transmiten a su descendencia) cruzándolos con ejemplares silvestres (*wild type*). Para la identificación de alelos mutantes, se utilizaron dos métodos consecutivos. En primer lugar, se extrajo gDNA de embriones individuales de cada cruce y se amplificó parte del exón afectado utilizando una polimerasa de alta eficiencia (iProof High-Fidelity DNA Polymerase, *Bio-Rad*). Se utilizó la pareja de oligonucleótidos 7-8 (Tabla Suplementaria 1). El producto de PCR fue sometido al ensayo de la endonucleasa T7. Posteriormente, en aquellas muestras que resultaron positivas en dicho ensayo, se amplificó la misma secuencia empleando una polimerasa que deja extremos cohesivos incorporando una base de adenina al extremo del producto amplificado (Advantage 2 PCR Kit, *Clontech*). El producto de PCR fue clonado en plásmidos pSpark (pSpark TA Done cloning kit, *Canvax Biotech*), que fueron transformados en bacterias DH5α competentes en placas con ampicilina, IPTG y X-Gal. Se identificaron colonias positivas mediante PCR y secuenciación.

Los alelos mutantes obtenidos *mYAPΔ7pb* y *mYAPΔ11pb* fueron detectados mediante una combinación de PCR y la posterior digestión del producto amplificado con las enzimas de restricción MspI y TfiI (*New England Biolabs*), respectivamente.

3.2. Generación de línea mutante *yap1b* en medaka. Se diseñaron dos moléculas de sgRNA contra el exón 2 del gen *yap1b* (Ensembl ID: ENSORLG00000005573) utilizando la herramienta CRISPRScan, descrita en una publicación anterior (Moreno-Mateos et al. 2015). Para cada diana, se diseñó un oligo de 52 nucleótidos, estructurado en tres partes: la secuencia del promotor T7, seguido de la diana CRISPR reconocida por la Cas9 sin la secuencia PAM y de una secuencia constante de 15 nucleótidos. Estos oligos (oligonucleótidos 9 y 10 especificados en la Tabla Suplementaria 1) fueron hibridados con un oligonucleótido reverso universal de 80

nucleótidos para generar una molécula dúplex de DNA de 117 pb. Los productos de PCR fueron purificados por columna. Cada dúplex de DNA se utilizó como base para una transcripción in vitro en la que se empleó el kit MAXIscript T7 (*ThermoFisher Scientific*). Las moléculas de sgRNA transcritas fueron resuspendidas en 12 µL de agua libre de nucleasas.

Ambas moléculas de sgRNA fueron coinyectados junto con proteína Cas9 bajo las mismas condiciones que las empleadas para la generación del mutante de *yap1*. Los alelos mutantes obtenidos se identificaron mediante idénticos procedimientos a los empleados para dicho gen. El alelo mutante puede ser identificado fácilmente mediante una sencilla PCR (utilizando los oligos 11 y 12 indicados en la Tabla Suplementaria 1).

3.3. Generación de línea mutante *ctgfa* en medaka. Se diseñaron cuatro moléculas de sgRNA contra el exón 2 del gen *ctgfa* (Ensembl ID: ENSORLG00000018064) utilizando la herramienta CRISPRScan, siguiendo el mismo procedimiento que en el subapartado anterior (oligonucleótidos 13-16 en la Tabla Suplementaria 1). Las cuatro moléculas de sgRNA fueron coinyectadas junto con la proteína Cas9 bajo las condiciones estándar. El alelo mutante *mCtgfa* Δ 10+13pb obtenido puede ser identificado mediante una combinación de PCR (oligonucleótidos 17-18 en la Tabla Suplementaria 1) y la posterior digestión del producto amplificado utilizando la enzima de restricción BanI (*New England Biolabs*).

4. iDamID-seq

4.1. Preparación de librerías. Para llevar a cabo el experimento de iDamID-seq y el posterior análisis de datos bioinformáticos, se siguió expresamente un protocolo desarrollado previamente para estudios en embriones de pez medaka (Gutiérrez-Triana et al. 2016). Se utilizó como base el plásmido pCS2+::cMyc-oDam-FL-GFP. La secuencia de GFP fue reemplazada por las secuencias codificantes de *yap1* (Ensembl ID: ENSDARG00000068401) y *wwtr1* (Ensembl ID: ENSDARG00000067719) de pez cebra, y de *yap1* y *yap1b* de medaka mediante una combinación de PCR utilizando dianas de restricción (oligos 19-26 en Tabla Suplementaria 1) y de ligación en el vector de destino. Las moléculas de mRNA de los cuatro cassettes de expresión fueron sintetizados utilizando el kit mMessage mMachine SP6 (*Ambion*) e inyectadas en embriones en estadio de una célula (Figura 5A) a una concentración de 10 ng/µL para el caso de los constructos de medaka y de 50 ng/µL para el caso de los constructos

de pez cebra. El grado de toxicidad de dichas concentraciones se testó con anterioridad y se concluyó que no suponían riesgo alguno para la viabilidad del embrión.

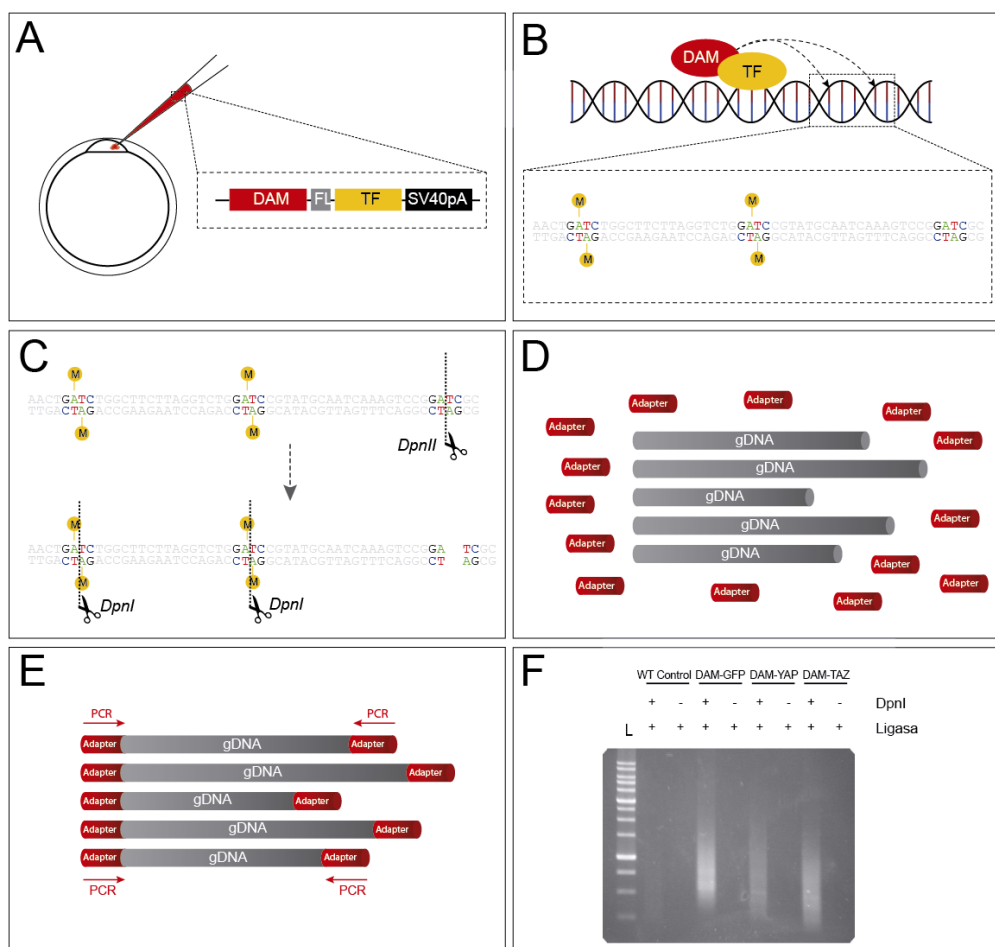


Figura 5 – Esquema del protocolo experimental de la técnica iDamID-seq. **A)** Los plásmidos con los constructos DAM-FT son inyectados en embriones de pez cebra y medaka en estadio de una célula. **B)** Cuando el FT se une a su diana transcripcional, la proteína DAM metila los sitios GATC adyacentes. **C)** Tras extraer la fracción de gDNA de los embriones inyectados, el genoma se fragmenta primero con la enzima *DpnII* y luego se cortan los sitios GATC metilados empleando la enzima *DpnI*. **D)** Se incorporan adaptadores a ambos extremos. **E)** Se amplifica cada fragmento desde sus adaptadores para generar librerías. **F)** Las librerías construidas correctamente generan una señal de barrido en la electroforesis de comprobación.

Se extrajo gDNA de embriones inyectados de pez cebra en estadio de 80% epibolia y de embriones inyectados de pez medaka en estadio 16. Para cada una de las tres réplicas biológicas preparadas por muestra inyectada, se empleó un total de 40 embriones. El gDNA extraído fue fragmentado utilizando la enzima de restricción *DpnII* (*NEB*), y posteriormente desfosforilado mediante tratamiento con fosfatasa alcalina (*NEB*). Estas muestras fueron digeridas posteriormente con la enzima *DpnI* (*NEB*) que reconoce específicamente secuencias GATC metiladas por la proteína DAM (Figuras 5B y 5C). A cada fragmento de DNA se le incorporó a ambos extremos un adaptador (Figura 5D) generado mediante hibridación de dos

oligos (oligos 27 y 28 de la Tabla Suplementaria 1). Acto seguido, los fragmentos de DNA fueron amplificados en una PCR de 25 ciclos (Figura 5E) utilizando oligos reforzados con enlaces fósforotioato (oligo 29 de la Tabla Suplementaria 1). Cada producto amplificado fue examinado en una electroforesis en gel de agarosa al 1% para confirmar la presencia de un barrido en las muestras tratadas con DpnI, en contraste con muestras control, no tratadas con DpnI (Figura 5F). Tras comprobar el éxito de la PCR, ésta fue repetida por quintuplicado para cada muestra con el fin de disponer de material suficiente para su posterior secuenciación. Se eliminó el exceso de DNA molde mediante un tratamiento con exonucleasa T7. Finalmente, las librerías fueron generadas según el protocolo citado anteriormente y secuenciadas en un equipo NGS Illumina HiSeq 2500.

4.2. Alineamiento de secuencias. Los datos crudos de secuenciación se procesaron utilizando la herramienta *Cutadapt* descartando las lecturas con una longitud inferior a 34 nucleótidos (parámetro `-m 34`) y eliminando aquéllas correspondientes a los adaptadores utilizados durante el proceso de secuenciación. Las lecturas de pez cebra se mapearon contra la versión *GRCz10/danRer10* del genoma de este organismo y las de pez medaka contra la versión *NIG/UT MEDAKA1/oryLat2* de su genoma utilizando *Bowtie2*. Se suprimió toda presencia de secuencias repetidas empleando la herramienta *samtools rmdup*.

4.3. Identificación de secuencias enriquecidas. El conjunto de fragmentos *DpnI* potenciales de cada organismo se construyó utilizando como base el paquete *BSgenome* de la versión correspondiente del genoma de cada especie (*danRer10* y *OryLat2*), utilizando la función *vMatchPattern* en combinación con el sitio de restricción 'GATC'. Sólo se consideraron para el análisis fragmentos adyacentes a dichas dianas de restricción con longitudes comprendidas entre 200 y 2000 pares de bases. Posteriormente, todas las lecturas mapeadas contra los fragmentos predichos considerados para el análisis se contaron una a una utilizando la función de R *summarizeOverlaps*, activando el parámetro *ignore.strand*. Con el objeto de descartar falsos positivos, sólo se consideraron para el análisis los fragmentos con un mínimo número de lecturas normalizadas con respecto a su longitud. Concretamente, este límite fue establecido como tres veces el número de lecturas totales para el total de fragmentos considerado, dividido por su longitud total. Con el conjunto resultante de regiones genómicas obtenidas, el recuento de lecturas fue calculado de nuevo mediante la función de R *summarizeOverlaps*. Se calcularon diferencias significativas entre cada conjunto de muestras a analizar y sus

correspondiente conjuntos de picos control (picos DAM-GFP) aplicando el paquete de R *DESeq2* (Love et al., 2014) a las matrices obtenidas. Todas estas operaciones están implementadas en la última versión del paquete de R *iDEAR*, disponible en la dirección web <https://bitbucket.org/juanlmateo/idear>. Se utilizó la versión 3.4 de R.

4.4. Análisis de genes asociados. A cada población de picos se le asoció el gen más cercano utilizando el comando *bedtools closest*. La lista completa de genes de cada organismo utilizado se descargó de la base de datos de ENSEMBL (Versión 89) utilizando la herramienta Biomart (www.ensembl.org/biomart) y fue procesada de tal manera que se indicara para cada gen la posición de su sitio de inicio de la transcripción (TSS).

4.5. Asociación de picos DamID a regiones abiertas de cromatina. Para examinar el porcentaje de picos solapantes con regiones abiertas de cromatina, las coordenadas de dichos picos fueron comparadas con las de picos de ATAC-seq utilizando el comando *bedtools intersect*. Los datos de ATAC-seq de pez cebra, al estar mapeados contra una versión anterior del genoma (danRer7), tuvieron que ser convertidos a la versión actualizada empleada en el análisis de picos de DamID utilizando la herramienta *LiftOver*. El conjunto de picos solapantes con marcas de ATAC-seq fue separado en dos grupos, en función de si ambas marcas solapaban o no con el TSS de su gen asociado.

4.6. Análisis de motivos. Se utilizó la herramienta DREME (Bailey et al., 2011) para determinar la existencia de motivos predefinidos en el conjunto de picos específicos de las cuatro proteínas examinadas (parámetro -p) con respecto a sus correspondientes conjuntos de picos control DAM-GFP (parámetro -n). Para fortalecer el grado de significación estadística del análisis, las muestras de picos control fueron procesadas para que éstas tuvieran la misma distribución de tamaños que la del conjunto de picos específicos frente al que se pretendía comparar y se quintuplicó su número. Como análisis complementario, se empleó la herramienta AME (McLeay y Bailey, 2010) aplicando una prueba U no paramétrica de Mann-Whitney y empleando las mismas colecciones de picos con el objeto de determinar en cada subpoblación la existencia del motivo TEAD definido en la base de datos JASPAR (MA0090.1). Este mismo motivo fue utilizado para generar una nueva búsqueda de motivos empleando la herramienta FIMO (Grant et al., 2011). El conjunto de picos de cada grupo de datos fue

clasificado en orden decreciente por su puntuación obtenida en el alineamiento llevado a cabo mediante el programa iDEAR, expresado en porcentaje de significación. Para cada porcentaje, se determinó el ratio entre el número de secuencias con al menos un motivo FIMO positivo y un número determinado de secuencias al azar con al menos un motivo positivo. Estas secuencias al azar fueron obtenidas tras intercambiar la posición de parejas de nucleótidos de todas las secuencias de manera aleatoria.

4.7. Análisis de términos de ontología genética. Se llevaron a cabo pequeños análisis de ontología genética (GO) para determinar la función principal llevada a cabo por el conjunto de genes asociados a los picos identificados durante el estudio de datos iDamID-seq. Se utilizó el algoritmo *GO-Slim* de la herramienta PANTHER (Mi et al., 2013). Sólo se consideraron significativos aquellos términos con un valor FDR inferior a 0.05.

5. RNA-seq

5.1. Preparación de librerías. Se pulverizaron embriones individuales *wild type* y *yap1*^{-/-} de medaka en estadio 16 en presencia de TRIzol (Ambion). Las muestras se centrifugaron a máxima velocidad y se reservó la fase superior. A dicho sobrenadante se le incorporó cloroformo para separar la fracción de DNA de la de RNA. El DNA fue precipitado en presencia de glicógeno y etanol 100% incubando la muestra a temperatura ambiente durante 20 minutos y centrifugando posteriormente a velocidad máxima durante 30 minutos a 4°C. Tras tres pasos de lavado consecutivos con etanol 75%, el DNA precipitado de cada muestra fue resuspendido en 30 µL de buffer TE. Las fracciones de RNA se precipitaron añadiendo glicógeno de calidad RNA (ThermoFisher Scientific) e isopropanol y siguiendo el mismo protocolo de incubación a temperatura ambiente y centrifugado en frío que el aplicado para las muestras de DNA. Finalmente, tras dos pasos de lavado con etanol 75%, cada muestra de RNA fue resuspendida en 12 µL de agua libre de nucleasas. El genotipo de cada embrión fue identificado aplicando sobre cada fracción de DNA el protocolo descrito anteriormente para la identificación del alelo *yap1* mutante. Las muestras de RNA correspondientes a cada genotipo fueron agrupadas para generar tres réplicas biológicas, que fueron secuenciadas en un equipo Illumina HiSeq 2500 (Figura 6).

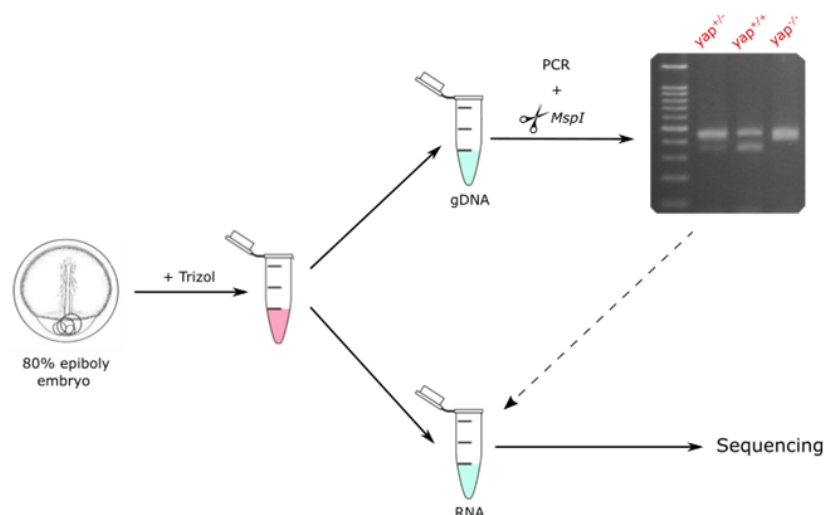


Figura 6 – Extracción de muestras de RNA y DNA para RNA-seq. Los embriones de pez medaka en estadio 16 (80% epibolia) son pulverizados en TRizol, del que se separan las fracciones de RNA y de gDNA de cada embrión. Se utiliza la fracción de gDNA para genotipar cada embrión individualmente (como ejemplo, se muestra el protocolo de genotipado de la línea mutante mYAPΔ7pb) y, en función del resultado, las muestras de RNA de cada genotipo se combinan para generar cada una de las réplicas enviadas a secuenciar.

5.2. Análisis de datos de RNA-seq. Las lecturas de RNA-seq obtenidas fueron mapeadas contra el genoma de medaka (oryLat2) utilizando el programa TopHat2, empleando los parámetros por defecto. Para determinar el nivel de expresión de cada gen para cada condición, se utilizó el programa Cufflinks y, con el objeto de determinar si existían diferencias de expresión significativas para cada caso entre la condición control y los embriones mutantes, se empleó la herramienta Cuffdiff. Sólo se consideraron como genes expresados diferencialmente aquéllos con una puntuación FDR (False Discovery Rate) inferior a 0.05.

6. Hibridaciones in situ

6.1. Preparación de sondas. Se generaron sondas para experimentos de hibridación in situ en embriones completos (*Whole-mount ISH*) amplificando, mediante PCR, preferentemente regiones de cDNA de embriones wild type de medaka de estadio 24, identificadas mediante el aislamiento de marcadores de secuencia expresada (EST, por sus siglas en inglés) de los genes *ntl*, *yap1* y *yap1b*. Las muestras de cDNA de estos embriones se obtuvieron primero mediante extracción de su mRNA siguiendo el protocolo descrito en el apartado anterior y posteriormente mediante retrotranscripción de estas moléculas de mRNA recurriendo al kit

iScript cDNA Synthesis (*BioRad*). Se utilizó la polimerasa de alta afinidad del kit Clontech Advantage 2 PCR (*Takara*) para generar productos amplificados con una base de adenina extra en sus extremos (se emplearon los oligos 30-35 indicados en la Tabla Suplementaria 1), que permitiera su inserción en plásmidos pSC-A-amp/kan de StrataClone (*Agilent*). Cada sonda fue sintetizada utilizando la polimerasa T3 o la polimerasa T7 (*Roche*) dependiendo de la orientación del inserto, empleando en el proceso nucleótidos marcados con digoxigenina (Digoxigenin-11-UTP, *Roche*). Las ISH fueron llevadas a cabo sobre embriones de medaka de estadio 16 y 19 (en el caso de las in situs de *ntl*, se utilizaron embriones *yap1*^{+/+} y *yap1*^{-/-}), fijados y decorionados en PFA 4%, siguiendo el protocolo de Thisse et al. (2008). Las muestras fueron permeabilizadas con proteinasa K durante un periodo de tiempo máximo de 4 minutos. Las sondas sintetizadas previamente se utilizaron a una concentración final aproximada de 1 ng/μL. Durante los pasos previos a la incubación con el anticuerpo de anti-digoxigenina, los embriones fueron lavados con 0.05xSSC/PBS en concentraciones decrecientes de sales. Finalizada la ISH, los embriones procesados fueron observados y fotografiados en un estereoscopio (Leica).

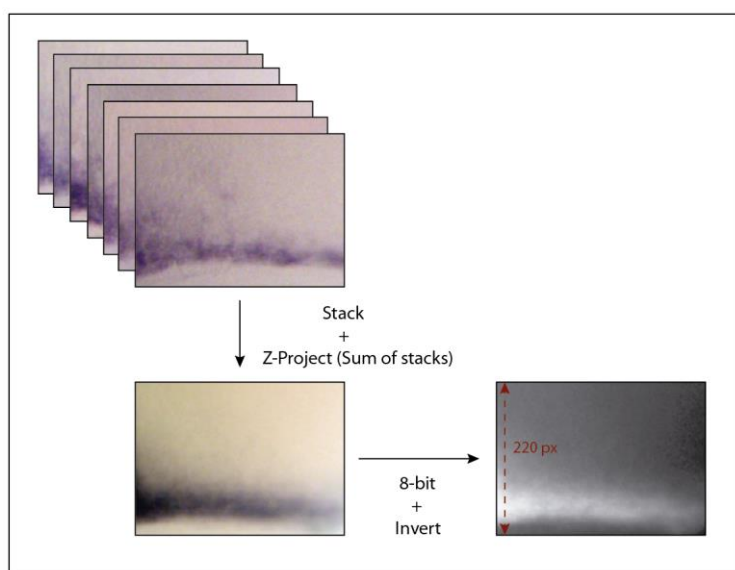


Figura 7 – Procesamiento de imágenes de hibridaciones in situ (ISH). Las imágenes tomadas correspondientes a un mismo genotipo son apiladas y transformadas para generar una imagen de 8-bits invertida cuya intensidad es medida a lo largo de una sección vertical cercana a la línea media.

VI.2. Análisis de imágenes. Las imágenes adquiridas de la sonda *ntl* sobre embriones *yap1*^{+/+} y *yap1*^{-/-} en estadio 16 y estadio 19 fueron procesadas con el software FIJI/ImageJ (versión 1.50i). Para el análisis de las imágenes de estadio 16, se tomaron secciones proximales a la línea media de cada embrión de 300 x 220 px de tamaño. Para analizar la distribución de

señales, las imágenes de cada genotipo fueron apiladas una encima de la otra y, tras convertirlas a una imagen de 8 bits e invertir sus colores, se cuantificó la intensidad de cada píxel a lo largo de toda la sección vertical adyacente a la línea media (Figura 7).

7. Análisis de expresión de genes de la familia Yap

7.1. RT-qPCR. Se obtuvo cDNA de embriones de medaka de estadios 7, 16 y 24 y de embriones de pez cebra de 3, 8 y 24 hpf siguiendo el protocolo de extracción de RNA descrito anteriormente. Se procedió a la cuantificación del nivel de expresión de *yap1* y de *yap1b* en medaka y de los genes *yap1* y *taz* en pez cebra mediante RT-qPCR (CFX96 Touch Real-Time PCR Detection System). Dichos niveles de expresión fueron normalizados primero con respecto al nivel de expresión del gen constitutivo *ef1a* y posteriormente con respecto al nivel de expresión más bajo registrado para cada gen (se emplearon los oligonucleótidos 36-47 especificados en la Tabla Suplementaria 1).

7.2. ddPCR. Para confirmar los datos obtenidos en la RT-PCR, se llevó a cabo una segunda PCR en tiempo real utilizando el sistema QX200™ Droplet Digital™ PCR (*Bio-Rad*). Se prepararon reacciones por duplicado de cada uno de los genes examinados junto con sus respectivos controles constitutivos a un volumen final de 22 µL utilizando 15 ng de cDNA por reacción, agua, la mezcla de oligonucleótidos utilizada para cada gen a una concentración final de 150 nM y el buffer de reacción QX200™ ddPCR™ EvaGreen Supermix (*Bio-Rad*), siguiendo las especificaciones de la casa comercial. Para excluir falsos positivos indicativos de una posible amplificación inespecífica, se prepararon controles negativos (NTC) para todas las parejas de oligonucleótidos examinadas utilizando agua en lugar de cDNA. Cada una de las reacciones de PCR preparadas fueron atomizadas en un conjunto de aproximadamente 22000 gotas de 1 nL utilizando el equipo QX200 Droplet Generator, y éstas fueron transferidas a placas de 96 pocillos (#12001925, *Bio-Rad*) que fueron termoselladas. Tras completarse la PCR, la cantidad relativa de cada producto amplificado se cuantificó en el equipo QX200 Droplet Reader examinando el grado de fluorescencia de cada microgota. Las lecturas obtenidas para cada gen y estadio fueron normalizadas con respecto a las del gen constitutivo correspondiente.

8. Ensayos funcionales

8.1. Inyección de morfolinos. Para reducir los niveles endógenos de las proteínas CTGFA y CYR61 en medaka, se diseñaron morfolinos de splicing (mCtgefMO and mCyr61MO), proporcionados por Gene Tools. El morfolino mCtgefMO-I1E2 (5' CAGCCTGAAATGATGGAGACACAGA) bloquea el sitio aceptor de splicing entre el intrón 1 y el exón 2 de la proteína CTGF (Figura 8A), mientras que el morfolino mCyr61MO-E2I2 (5' GCTGCATGGGCTGTATACTAACCTC) bloquea el aceptor de splicing situado entre el exón 2 y el intrón 2 de la proteína CYR61 (Figura 8B).

Ambos morfolinos se coinyectaron tanto por separado como conjuntamente a una concentración final de 100 μ M, junto con mRNA de LynTdTomato a una concentración final de 25 ng/ μ L como marcador, en embriones de medaka en estadio de una célula. Como control, se inyectó en embriones de la misma serie mRNA de LynTdTomato a una concentración final de 25 ng/ μ L. Para demostrar la eficiencia de ambos morfolinos, se obtuvieron muestras de cDNA de embriones de estadio 24, que fueron amplificadas por PCR (oligos 48-51 de la Tabla Suplementaria 1) para detectar la presencia de intrones retenidos o bien exones suprimidos en la secuencia.

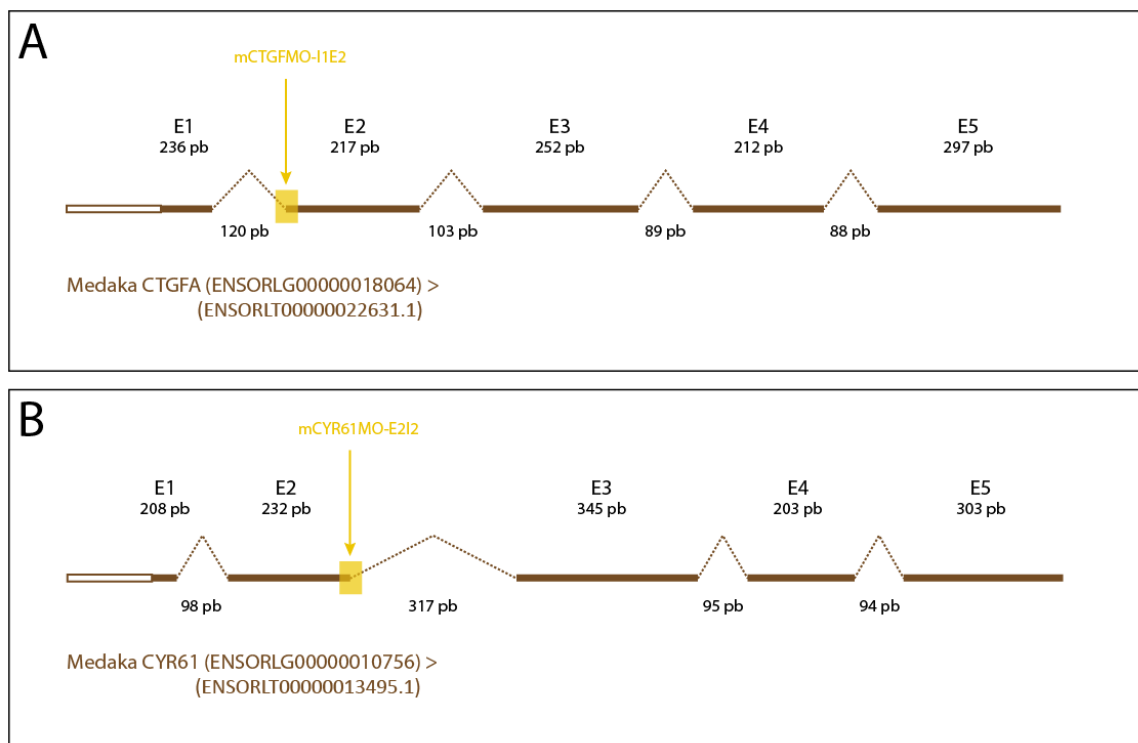


Figura 8 – Esquema de diseño de morfolinos. **A)** El morfolino mCTGFMO-I1E2 elimina el sitio aceptor de splicing situado entre el intrón 1 y el exón 2 del gen *ctgfa*. **B)** El morfolino mCYR61MO-E2I2 neutraliza el sitio de splicing ubicado entre el exón 2 y el intrón 2 del gen *cyr61*.

8.2. Sobreexpresión asimétrica. Se amplificó la secuencia de la región codificante del gen *ctgfa* (Ensembl ID: ENSDARG00000042934) de pez cebra (oligos 52-53 de la Tabla Suplementaria 1), que fue clonada en plásmidos pCS2+. Una vez confirmado el éxito de la ligación por secuenciación, este plásmido fue linealizado con la enzima NotI y se procedió a la transcripción in vitro de su inserto (SP6 mMessage mMachine Kit, *Ambion*). Se inyectó mRNA de *zCtgfa*, además de mRNA de LynTdTomato como marcador en una sola célula en embriones de pez medaka en estadio de 2, 4 y 8 células. Todas las muestras fueron inyectadas a una concentración final de 25 ng/μL.

9. Análisis de trayectorias celulares durante la gastrulación

9.1. Preparación de embriones. Se inyectaron embriones de medaka en estadio de una célula con una combinación de mRNA de un marcador nuclear verde (pCS-H2B-eGFP, *Addgene* #53744) y otro de membrana (LynTdTomato), ambos a una concentración final de 25 ng/μL. Los embriones fueron incubados durante tres horas a 28°C y posteriormente a 25°C hasta la mañana siguiente. Los embriones que mostraban una señal fluorescente óptima y distribuida homogéneamente al visualizarlos en el estereoscopio fueron seleccionados para proceder a su decorionado. Para ello, se siguió un proceso de tres fases: en primer lugar, los embriones fueron pasados por papel de lija (tamaño de grano 150) para eliminar restos de suciedad en el corion y debilitar su estructura externa; posteriormente, los embriones fueron incubados a 28°C en presencia de pronasa (20 mg/mL) durante 40 minutos para eliminar la capa externa del corion; finalmente, tras eliminar los restos de pronasa lavando los embriones cinco veces consecutivas con medio E3, éstos fueron incubados durante aproximadamente una hora a 28°C en presencia de *hatching enzyme* para debilitar la capa interna del corion. Los restos de corion fueron eliminados en presencia de medio BSS 1X (0.65% NaCl, 0.04% KCl, 0.02% MgSO₄·7H₂O, 0.02% CaCl₂·2H₂O y 0.001% Phenol red, esterilizado y ajustado a pH 8.3 con 5% NaHCO₃) utilizando unas pinzas afiladas. Los embriones decorionados fueron incubados a temperatura ambiente durante una hora en presencia de 1-heptanol a una concentración final de 3.5 mM en medio BSS. Este agente desacopla las uniones gap de las células del embrión en desarrollo para reducir los movimientos rítmicos contráctiles que tienen lugar en el blastodermo durante el desarrollo temprano del pez medaka y que de otra forma imposibilitarían una correcta visualización de la etapa de gastrulación. Los embriones seleccionados finalmente fueron colocados en placas de cultivo FluoroDish de 35 mm (*WPI*,

FD3510-100) dentro de una base de agarosa de bajo punto de fusión diluida al 0.5% en medio BSS con 1-heptanol.

9.2. Adquisición de imágenes. Todas las películas se grabaron en un microscopio confocal Leica SP5 utilizando un objetivo multi-inmersión 20X. En función del tipo de datos recabados, se utilizó una configuración diferente.

- a) Películas generales de gastrulación: se tomaron imágenes de una superficie de $775\ \mu\text{m}$ x $775\ \mu\text{m}$ (a una resolución de $0.757\ \mu\text{m}/\text{px}$) cada minuto durante un periodo de tiempo de 14 horas. Se aplicó un Z-stack de seis pasos para la toma de imágenes de $5\ \mu\text{m}$ de grosor en una sección de aproximadamente $30\ \mu\text{m}$ en el eje Z del embrión. Se emplearon los láseres de emisión de $488\ \text{nm}$ y de $561\ \text{nm}$ de longitud de onda. Para reducir al máximo el ruido de fondo de cada imagen, se emplearon detectores híbridos y se estableció la toma de tres imágenes consecutivas por cada línea de píxeles (Parámetro *Line average* establecido en el valor 3).
- b) Películas detalladas de gastrulación: se adquirieron imágenes de una superficie de $193.75\ \mu\text{m}$ x $193.75\ \mu\text{m}$ (a una resolución de $0.189\ \mu\text{m}/\text{px}$) cada minuto durante un periodo máximo de 120 minutos. Se aplicó un Z-stack de tres pasos para la toma de imágenes de aproximadamente $3\ \mu\text{m}$ de grosor en una sección de $10\ \mu\text{m}$ en el eje Z. Se emplearon los láseres de emisión de $488\ \text{nm}$ y de $561\ \text{nm}$ de longitud de onda. Para reducir al máximo el ruido de fondo de cada imagen, se emplearon detectores híbridos y se estableció la toma de tres imágenes consecutivas por cada línea de píxeles (Parámetro *Line average* establecido en el valor 3). Se tomaron secciones proximales a la línea media, tanto en el frente de epibolia como en el extremo anterior del embrión de manera simultánea utilizando el escáner de multi-posición del microscopio.

9.3. Procesamiento y análisis de imágenes. Las imágenes adquiridas fueron procesadas con el software FIJI/ImageJ. Se obtuvieron proyecciones máximas de los dos stacks más superficiales para cada conjunto de datos con el objeto de condensar en una sola imagen la intensidad máxima de cada píxel del stack especificado. Se equilibró el nivel de brillo y contraste para aumentar la nitidez de las imágenes y reducir el nivel de ruido de fondo.

Para el análisis de trayectorias celulares, se recurrió al plugin *Manual tracking* instalado por defecto en ImageJ. Se establecieron los parámetros de 60 segundos como intervalo de tiempo, el valor de 0.189 μm para el parámetro de calibración del eje X/Y y se aplicaron los valores por defecto del resto. Para determinar la velocidad y el grado de persistencia celular en cada condición experimental, se recurrió al programa DiPer, compuesto de un conjunto de macros ejecutables en Excel (Gorelik y Gautreau, 2014). Entre las medidas útiles que determinan el grado de persistencia de las células analizadas se encuentra el cociente o ratio de direccionalidad (d/D), que se define como el cociente entre la distancia más corta entre el punto de inicio y el punto final de desplazamiento de una célula (d) y la trayectoria real que sigue la célula entre esos dos puntos (D). Por tanto, cuanto más recta sea la trayectoria real de la célula, más alto será su ratio de direccionalidad. Otra de estas medidas es el desplazamiento cuadrático medio (MSD). En el contexto tisular, se definiría como la medida de la desviación de la posición de una célula con respecto a una posición de referencia en base al tiempo, es decir, es un valor que estima la superficie explorada por una célula a lo largo de un periodo de tiempo determinado. A partir de este valor se puede deducir si una determinada célula tiende a migrar por efecto de un agente quimioatrayente o bien lo hace mediante simple difusión pasiva. Cuanto más alto sea, mayor distancia habrá migrado una célula a una velocidad determinada. Por último, la velocidad media de cada célula constituye un parámetro importante para caracterizar su motilidad.

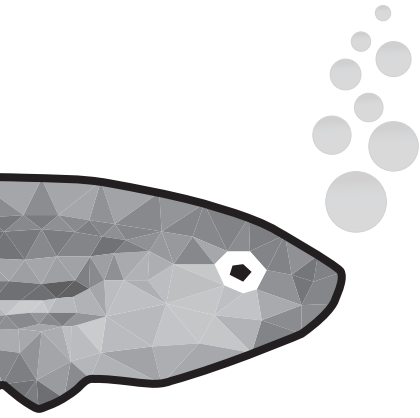
10. Análisis del grado de apoptosis y de proliferación celular en embriones mutantes

10.1. Inmunofluorescencia. Se colectaron embriones de un cruce entre ejemplares *yap1^{+/-}* *yap1b^{+/-}* que fueron fijados en PFA 4% en estadio 20 (5 somitas). Posteriormente, fueron decorionados con fórceps y, tras sucesivos pasos de lavado con PBS-Tween 0.2% (PBT) y un tratamiento con una solución de bloqueo (preparada con suero fetal bovino al 10% en volumen) a temperatura ambiente durante 2 horas, fueron incubados toda la noche a 4°C en presencia del anticuerpo primario correspondiente, en dilución 1:500. Así, para determinar el grado de muerte celular de los embriones, éstos fueron incubados en presencia de un anticuerpo anti-caspasa 3 de conejo (Purified rabbit anti-active caspase-3, *BD Biosciences*) y para analizar su grado de proliferación celular, se incubaron con anticuerpos anti-fosfohistona H3 (Anti-phospho-Histone H3, *Merck Millipore*). Al día siguiente, los embriones procesados fueron sometidos a nuevos periodos de lavado con PBT y fueron incubados en dilución 1:1000

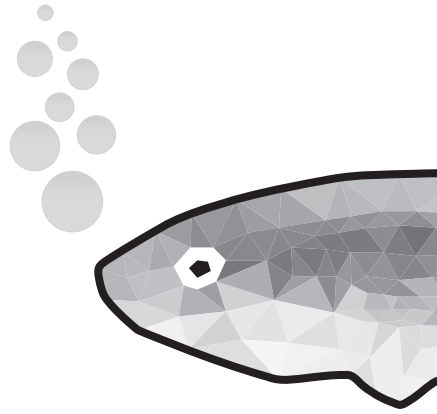
en oscuridad a 4°C con el anticuerpo secundario Alexa Fluor™ 555 Goat anti-rabbit (*Invitrogen*) durante toda la noche. Posteriormente, los núcleos de las células fueron teñidos tratando los embriones con DAPI (*Sigma*) en dilución 1:5000, y éstos fueron sometidos a un último paso de lavado con PBT. Los embriones incubados en presencia del anticuerpo primario anti-caspasa 3 fueron fotografiados en el estereoscopio.

10.2. Adquisición de imágenes. Los embriones seleccionados fueron montados en placas de cultivo FluoroDish, de manera idéntica a la mencionada en el apartado IX. Los embriones procesados fueron examinados en un microscopio de superresolución (Zeiss LSM 880) utilizando un objetivo de multi-inmersión 40x y se adquirieron imágenes de una superficie de 354.25 μm x 354.25 μm (a una resolución de 0.692 $\mu\text{m}/\text{px}$). Se aplicó un Z-Stack de 24 pasos para la toma de imágenes de una sección del embrión de 96 μm de grosor. Se emplearon los láseres de emisión de 405 nm y 561 nm de longitud de onda.

10.3. Análisis de imágenes. Todas las imágenes fueron procesadas en FIJI/ImageJ. Para cada una de las imágenes adquiridas en el microscopio, se generó la proyección máxima de cada canal y se redujo al mínimo el ruido de fondo (parámetro *Subtract Background* establecido en 200). Tras aplicar máscaras en ambos canales y un algoritmo de segmentación *Watershed* sobre la máscara del canal rojo, se identificaron las regiones marcadas con los anticuerpos que coincidieran con la posición de núcleos celulares y que ocuparan una superficie comprendida entre las 6 y las 200 μm^2 (para evitar seleccionar partículas correspondientes a restos que no pudieron ser eliminados satisfactoriamente durante los lavados del experimento de inmunofluorescencia). Se procedió al recuento de señales de muerte y proliferación celular tanto en el total de la imagen como exclusivamente en la superficie ocupada por el embrión, y se determinó la tasa de señales positivas por mm^2 tanto en el embrión como en el área correspondiente al vitelo.



Resultados



1. Identificación de Yap1b como nuevo miembro de la familia de proteínas Yap

En un artículo previo, y en el contexto de un estudio de mutagénesis química (ENU) a gran escala en el pez medaka (Furutani-Seiki et al., 2004), se identificó una mutación en el gen *yap1*, que genera un codón de stop prematuro en el primer dominio WW de la proteína. Estos mutantes, denominados *hirame* (*hir*), muestran profundas anomalías anatómicas, entre las que destacan un plegamiento muy deficiente de la copa óptica, la desalineación de la lente con respecto a ésta, un evidente retraso en el cierre del blastoporo y un incremento en la tasa de muerte celular durante las etapas más tempranas de su desarrollo (Porazinski et al., 2015). Nuestra investigación confirmó estos resultados generando dos alelos mutantes adicionales mediante la tecnología CRISPR/Cas9 en medaka. Las lesiones cromosómicas obtenidas consisten en sendas deleciones de 7 y 11 pb en el exón 2 de la proteína, que en ambos casos producen un cambio en la pauta abierta de lectura de la secuencia codificante y, en consecuencia, la aparición de un codón de stop prematuro antes del primer dominio WW. En ambos casos, se generan versiones truncadas de la proteína (Figura Suplementaria 3). Ambos alelos muestran un fenotipo muy similar al descrito previamente en la línea mutante *hir*, con la excepción de que el defecto en el cierre del blastoporo parece más leve (Figura Suplementaria 3). El fenotipo severo observado en mutantes *yap1* en medaka contrasta con los defectos mucho más leves descritos en embriones mutantes para *yap1* en pez cebra. De hecho, el fenotipo de los mutantes *yap1* de medaka más bien se asemeja a las anomalías anatómicas observadas en mutantes dobles *yap1/taz* de pez cebra (Agarwala et al., 2015; Kimelman et al., 2017; Miesfeld et al., 2015).

Con el fin de determinar si la discrepancia observada entre los fenotipos de los mutantes de pez cebra y medaka podía deberse a una contribución diferente entre los distintos parálogos, se llevó a cabo un análisis filogenético de todos los miembros de la familia Yap en teleósteos. Se construyó para ello un árbol filogenético mediante el alineamiento múltiple de secuencias de proteínas de esta familia entre 7 especies representativas de teleósteos. Como referencia externa, se incluyeron también en este análisis las secuencias de humano y ratón. A diferencia de la mayoría de estudios, que describen la existencia de dos parálogos, Yap1 y Taz, en todos los vertebrados, nuestro análisis filogenético muestra inequívocamente la existencia de no dos sino tres grupos de parálogos bien diferenciados, a saber: un primer grupo que incluye a todas las proteínas conocidas como *Yap1*, un segundo grupo que incluye a todas las proteínas *Taz*, y un tercer grupo, denominado en este proyecto de investigación como *Yap1b*, que muestra un grado de similitud algo mayor con Yap1 que con Taz (Figura 9A).

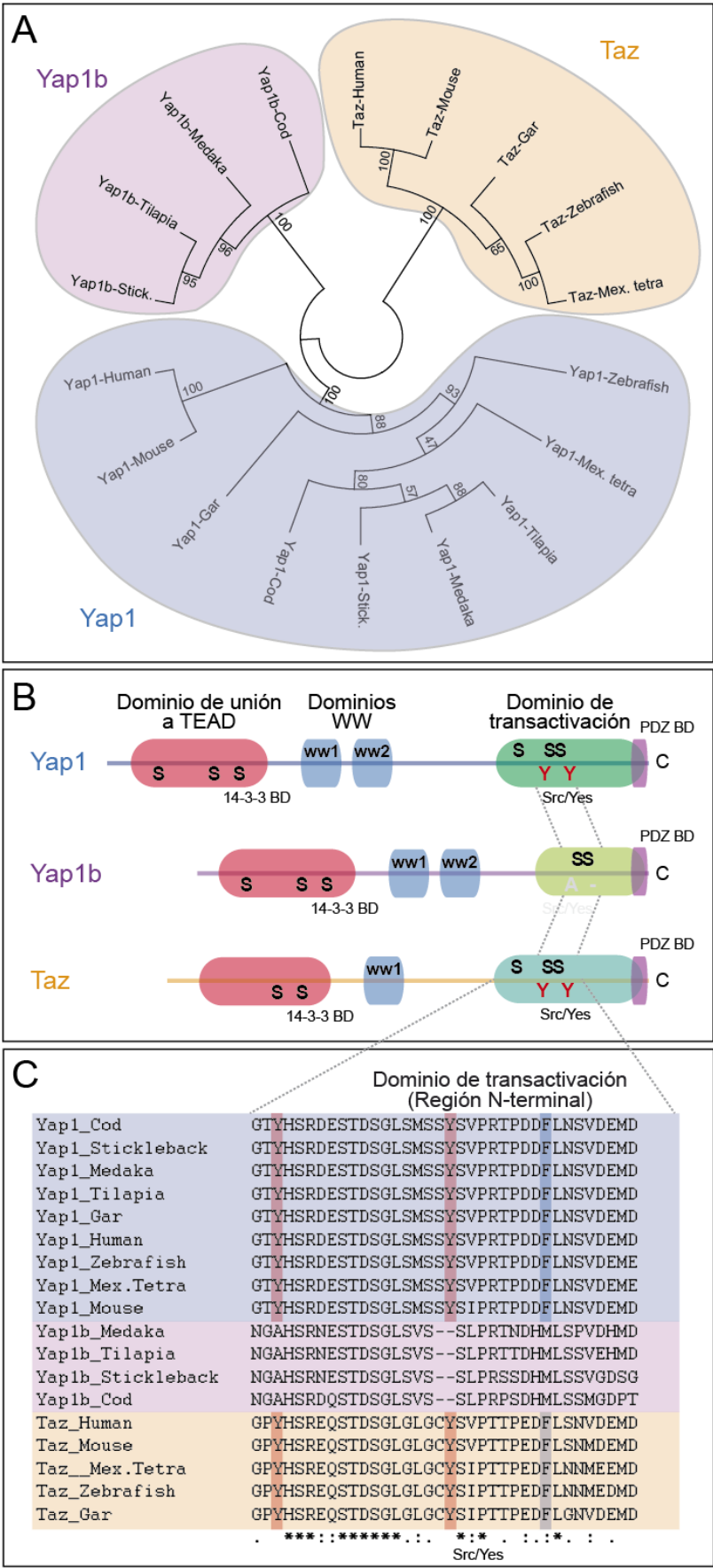


Figura 9 – Análisis filogenético de proteínas de la familia Yap en teleósteos. **A)** Los análisis de máxima verosimilitud llevados a cabo determinan la existencia de tres subfamilias de proteínas Yap: Yap1, Yap1b y Taz. **B)** Representación esquemática de los diferentes dominios que conforman las proteínas Yap1, Yap1b y Taz. Obsérvese la estructura divergente del dominio de transactivación carboxi-terminal de Yap1b. **C)** Los alineamientos múltiples de secuencia de la región amino-terminal del dominio de transactivación muestran las diferencias encontradas en la secuencia de Yap1b. Nótese la ausencia o sustitución de residuos de tirosina (rojo) y de fenilalanina (azul).

Se pudo identificar la presencia de genes que codifican para proteínas Yap1b en todas las especies del clado *Euteleosteomorpha*, el más grande en número de especies de entre los teleósteos (Betancur et al., 2017), que incluye especies como el bacalao común (*Gadus morhua*), el pez medaka (*Oryzias latipes*), el pez tilapia (*Oreochromis niloticus*) y los peces espinosos (*Gasterosteus aculeatus*). Esta observación sugiere la posibilidad de que el gen *yap1b* surgiera tras la duplicación genómica específica de los teleósteos (Amores et al., 1998). Para comprobar esta hipótesis, se examinó el grado de sintenia de la región genómica que contiene los genes de la familia Yap en teleósteos. Los resultados de este análisis indican que los loci que contienen a los genes *yap1* y *yap1b* derivan de un ancestro común tras la duplicación del genoma de teleósteos, mientras que los genes cercanos a *taz* muestran una disposición cromosómica independiente (Figura Suplementaria 4). Estos resultados confirman que, a diferencia de las hipótesis planteadas en estudios previos (Porazinski et al., 2015), el segundo parálogo de la familia Yap presente en el pez medaka no es *taz* sino *yap1b*.

Se procedió entonces a analizar en detalle la estructura de las proteínas Yap1b. El alineamiento múltiple de secuencias mostró que los dominios n-terminal Tead y WW de Yap1b están relativamente bien conservados. Concretamente, el dominio de unión a Tead de esta proteína presenta una mayor similitud con respecto a su equivalente en Yap1 (65.5%) que respecto a Taz (40.3%) (Figura Suplementaria 5). Sin embargo, el dominio de activación transcripcional de Yap1b parece estar mucho menos conservado con respecto a los dominios de transactivación de Yap1 y de Taz (porcentajes de identidad del 44.8% y del 23.4%, respectivamente). Este contraste en el porcentaje de similitud entre dominios encierra un dato interesante: todos los aminoácidos de tirosina presentes en el dominio de activación transcripcional de Yap1b, fosforilables por kinasas de la familia Src, así como los residuos de fenilalanina, que son estructuralmente muy parecidos, han sido sustituidos por otros aminoácidos conservados o directamente han sido suprimidos (Figura 9B-C; Figura Suplementaria 5D). Estas modificaciones podrían alterar radicalmente las propiedades reguladoras de esta proteína, puesto que la fosforilación de este dominio por kinasas tipo Src tiene un papel clave en la formación y activación transcripcional de los complejos Yap/Tead (Li et al., 2016).

Para comprender la historia evolutiva de las proteínas de la familia Yap, particularmente de la evolución de Yap1b, se profundizó en el análisis genómico de cuatro especies adicionales que pertenecen a ramas clave del árbol evolutivo de los teleósteos (Figura suplementaria 6): la anguila japonesa (*Anguilla japonica*, del superorden *Elopomorpha*), el pez elefante (*Paramormyrops kingsleyae*, del superorden *Osteoglossomorpha*), el arenque común (*Clupea*

harengus, de la cohorte *Otomorpha*, perteneciente al orden de los *Cupleiformes*), y el lucio europeo (*Esox lucius*, de la cohorte de los *Euteleosteomorpha*, perteneciente al superorden *Protacanthopterygii*). El análisis de secuencia múltiple del dominio de activación transcripcional de las proteínas de la familia Yap de estas especies revela la existencia de cuatro parálogos en los teleósteos más ancestrales (la anguila japonesa y el pez elefante), que representan el resultado de la duplicación de los genes *yap1* y *taz* originales presentes en otros vertebrados (Tabla Suplementaria 2; Figura Suplementaria 6). Tomando como referencia la conservación general de las secuencias proteicas, en especial de algunos residuos clave, fue posible identificar la presencia tanto de *yap1* como de *yap1b* en el genoma de los teleósteos menos evolucionados. No obstante, el dominio de activación transcripcional de *yap1b* parece ser menos divergente con respecto al de *yap1* en estas especies, y sus residuos de tirosina y fenilalanina se encuentran al menos parcialmente conservados (Figura Suplementaria 6A). Estos análisis son consistentes con la pérdida evolutiva de *yap1b* en toda la cohorte de los *Otomorpha*, que consta de más de 8000 especies vivas en la actualidad, incluyendo algunas como el arenque, el tetra mexicano (*Astyanax mexicanus*) y el pez cebra. También sugieren una divergencia evolutiva en el dominio de transactivación de Yap1b poco después de la aparición de la cohorte de los *Euteleosteomorpha*, que incluye especies como el lucio, el bacalao, el pez medaka, el pez tilapia y el grupo de peces espinosos. Finalmente, a pesar de que la historia evolutiva de los dos parálogos de *taz* parece más compleja y que su comprensión requerirá de investigaciones más exhaustivas, parece que ambas copias se perdieron consecutivamente a lo largo de la evolución de los teleósteos. Así, sólo una copia está presente en el grupo de los *Otomorpha*, y ninguna en el clado de los *Percomorpha*, que cuenta con aproximadamente 15000 especies vivas (lo que constituye la cuarta parte de todas las especies de vertebrados) (Figura Suplementaria 6B). En resumen, los análisis filogenéticos llevados a cabo en el presente estudio revelan la compleja historia evolutiva de las proteínas de la familia Yap en los teleósteos, e identifican el origen evolutivo de *yap1b*, una proteína paróloga con un dominio de activación transcripcional divergente que desplaza a *taz* como parálogo de *yap1* en las especies del clado de los *Euteleosteomorpha*.

2. Análisis de expresión de *yap1* y *yap1b* durante el desarrollo temprano

Para entender el papel desempeñado por *yap1b* durante las primeras etapas de la embriogénesis de medaka, nuestros análisis se enfocaron a la ventana de desarrollo comprendida entre las postrimerías de la gastrulación (estadios 15-16, cuando la epibolia ha

alcanzado el 80% de la superficie del embrión) y la formación de las copas ópticas (estadios 23-24). Este periodo de desarrollo abarca las etapas en las que se aprecian por primera vez los defectos morfogénéticos identificados en embriones *yap1*^{-/-} de medaka (Porazinski et al., 2015) y en embriones *yap1*^{-/-}; *taz*^{-/-} de pez cebra (Kimelman et al., 2017). Estudios previos llevados a cabo en embriones de pez cebra muestran una distribución bastante amplia en el patrón de expresión de *yap1*, que es especialmente acusada en la placa neural y en el mesodermo axial (Kimelman et al., 2017; Thisse et al., 2008). Para extrapolar estas observaciones a embriones de medaka, se investigó el patrón de expresión de *yap1* y de *yap1b* mediante hibridaciones in situ (ISH) y PCR cuantitativas (qPCR) (Figura 10).

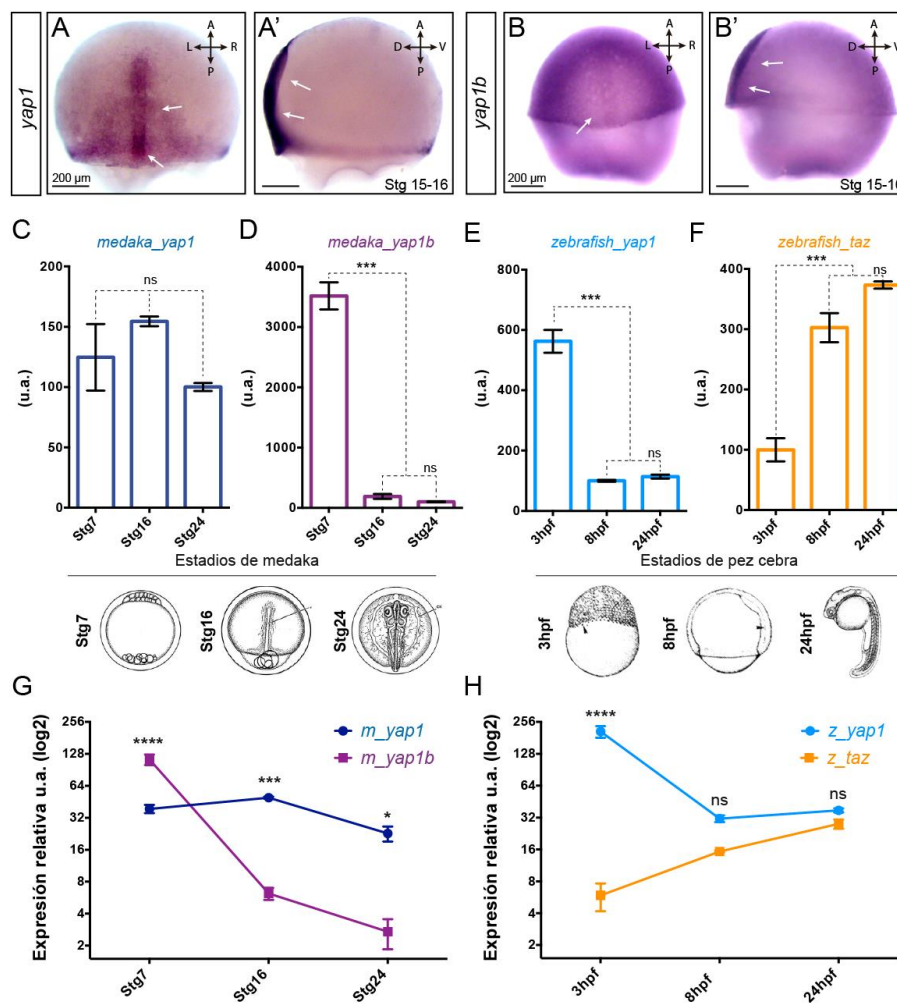


Figura 10 – Análisis de expresión de genes de la familia Yap. Análisis del patrón de expresión de los genes *yap1* (A, A') y *yap1b* (B, B') mediante hibridación in situ (ISH) en embriones de medaka de estadio 15-16. Escala: 200 µm. Cuantificación (en unidades arbitrarias) de los niveles de expresión de genes de la familia Yap a lo largo de tres etapas de desarrollo equivalentes (mórula temprana, gastrulación tardía y fase de formación de la copa óptica) en embriones de medaka (C, D) y pez cebra (E, F) mediante qRT-PCR (n = 3). Los niveles de expresión relativa de estos parálogos (en unidades arbitrarias) también fueron determinados mediante ddPCR (n = 2) en embriones de medaka (G) y pez cebra (H). Se realizó un análisis ANOVA (One-way ANOVA) seguido de un test de Fisher entre las diferentes condiciones para determinar el grado de significación en el nivel de expresión observado.

Los análisis de ISH durante la etapa de 80% epibolia determinaron que ambos genes muestran en medaka un patrón de expresión similar al del ortólogo de *yap1* en pez cebra. Ambos genes se expresan casi ubicuamente, aunque están particularmente enriquecidos en el mesodermo axial (Figura 10A-B). Se examinaron entonces los niveles de transcripción relativa de *yap1* y *yap1b* mediante qPCR en tres etapas de desarrollo diferentes: antes de la activación de la transcripción cigótica (estadio 7), durante la gastrulación tardía (estadio 16) y durante la fase de formación de las copas ópticas (estadio 24). Este análisis revela que, mientras que los niveles de *yap1* se mantienen relativamente constantes en medaka, los de *yap1b* presentan una enorme contribución materna y su nivel de expresión desciende mucho tras la activación de la transcripción cigótica (Figura 10C-D). A fin de establecer un análisis comparativo, también se compararon los niveles de expresión de *yap1* y de *taz* en embriones de pez cebra en etapas de desarrollo equivalentes. Los resultados revelan que, a diferencia de *taz*, *yap1* presenta cierta contribución materna (Figura 10E-F). Para obtener una medición más precisa de los niveles absolutos de RNA de cada gen, se llevó a cabo un experimento de ddPCR (Digital Droplet PCR) (Pinheiro et al., 2012). Este método confirmó los resultados del experimento de qPCR y nos permitió estimar la contribución relativa de cada parólogo tanto en pez cebra como en medaka (Figura 10G-H). Estos datos indican que en los embriones de pez cebra, tanto *yap1* como *taz* se expresan a unos niveles similares tras la transición materna-cigótica (*maternal-to-zygotic transition*, MZT). Por el contrario, y a pesar de contar con una fuerte contribución por parte del genoma materno, el gen *yap1b* de medaka se expresa a niveles muy bajos tras la MZT. Esta observación, por sí misma, podría explicar el fuerte fenotipo observado tras la mutación de *yap1* en medaka.

3. Identificación del conjunto de sitios de unión a la cromatina de proteínas de la familia Yap durante la embriogénesis temprana en teleósteos mediante iDamID-seq

Para investigar las propiedades reguladoras de las proteínas de la familia Yap en teleósteos, especialmente las de *yap1b*, se procedió a la identificación de sus regiones diana en el genoma. Debido a la ausencia de anticuerpos aptos para llevar a cabo experimentos de ChIP-seq contra las proteínas de la familia Yap en teleósteos, se optó por utilizar la novedosa técnica iDamID-seq para determinar su perfil de unión al genoma. El método de identificación de la DNA adenina metiltransferasa (DamID) fue desarrollado originalmente en *Drosophila* (van Steensel et al., 2001), y se basa en la expresión de la secuencia de factores de transcripción y

otros modificadores de la cromatina, fusionada a la de la metilasa bacteriana DAM. A pesar de su amplio uso en *Drosophila* (Aughey y Southall, 2016), no ha sido hasta hace poco tiempo cuando ha sido adaptado para llevar a cabo estudios in vivo en vertebrados (Gutiérrez-Triana et al., 2016; Tosti et al., 2018). En este trabajo hemos seguido el procedimiento iDamID-seq/iDEAR, establecido originalmente en medaka (Gutiérrez-Triana et al., 2016). Se generaron constructos *DAM:yap1* y *DAM:yap1b* para medaka, y sus correspondientes transcritos, así como la molécula control *DAM:GFP* de mRNA, que fueron inyectados en embriones de medaka en estadio de una célula. Los embriones inyectados (aproximadamente 40 por cada condición experimental y réplica) fueron procesados en el estadio de 80% epibolia (estadio 16) para fijar el interactoma genético de ambos constructos establecido durante la gastrulación. Con el objetivo de establecer un análisis comparativo, estos experimentos de iDamID-seq también fueron llevados a cabo en embriones de 80% epibolia de pez cebra (8 hpf), a los que se inyectó mRNA de constructos *DAM:yap1*, *DAM:taz* y *DAM:GFP* en estadio de una sola célula. Tras aislar y amplificar la regiones metiladas por la proteína DAM en tres réplicas biológicas independientes, las muestras fueron enviadas para su secuenciación y las lecturas obtenidas fueron mapeadas sobre la última versión de los genomas de referencia de pez cebra y medaka. La correlación de lecturas mapeadas fue elevada entre réplicas (por encima del 70%) y, como se esperaba, muy baja al compararlas con sus respectivos controles *DAM:GFP* (Figura Suplementaria 7).

Se siguió el protocolo iDEAR (Gutiérrez-Triana et al., 2016) para identificar y comparar el patrón de picos metilados de cada proteína de la familia Yap con respecto a su control correspondiente. Como ejemplo representativo, se muestra en este apartado el conjunto de picos identificados cercanos al locus del gen *amotl2a*, una conocida diana transcripcional de Yap1, tanto para pez cebra como para medaka (Figura 11). En el material adicional de este documento se incluyen otros ejemplos adicionales para otros genes regulados por *yap1*, como *ctgfa*, *cyr61*, *lats2*, *ccne1*, *boka* y el propio *yap1* (Figura Suplementaria 8).

Se identificaron un total de 4212 y 3792 picos de DamID en el genoma de medaka para *yap1* y *yap1b*, y 7553 y 8608 picos en el genoma de pez cebra para los genes *yap1* y *taz*, respectivamente (Tabla 1; Tabla Suplementaria 3). En el caso de medaka, hallamos un importante solapamiento entre picos de *yap1* y picos de *yap1b* (aproximadamente el 65% de los picos de *yap1b* también eran positivos para *yap1*), mientras que sólo el 28.6% de los picos de *yap1* también resultan positivos para *taz* en pez cebra (Figura 11A-C; Figura Suplementaria 9A-B).

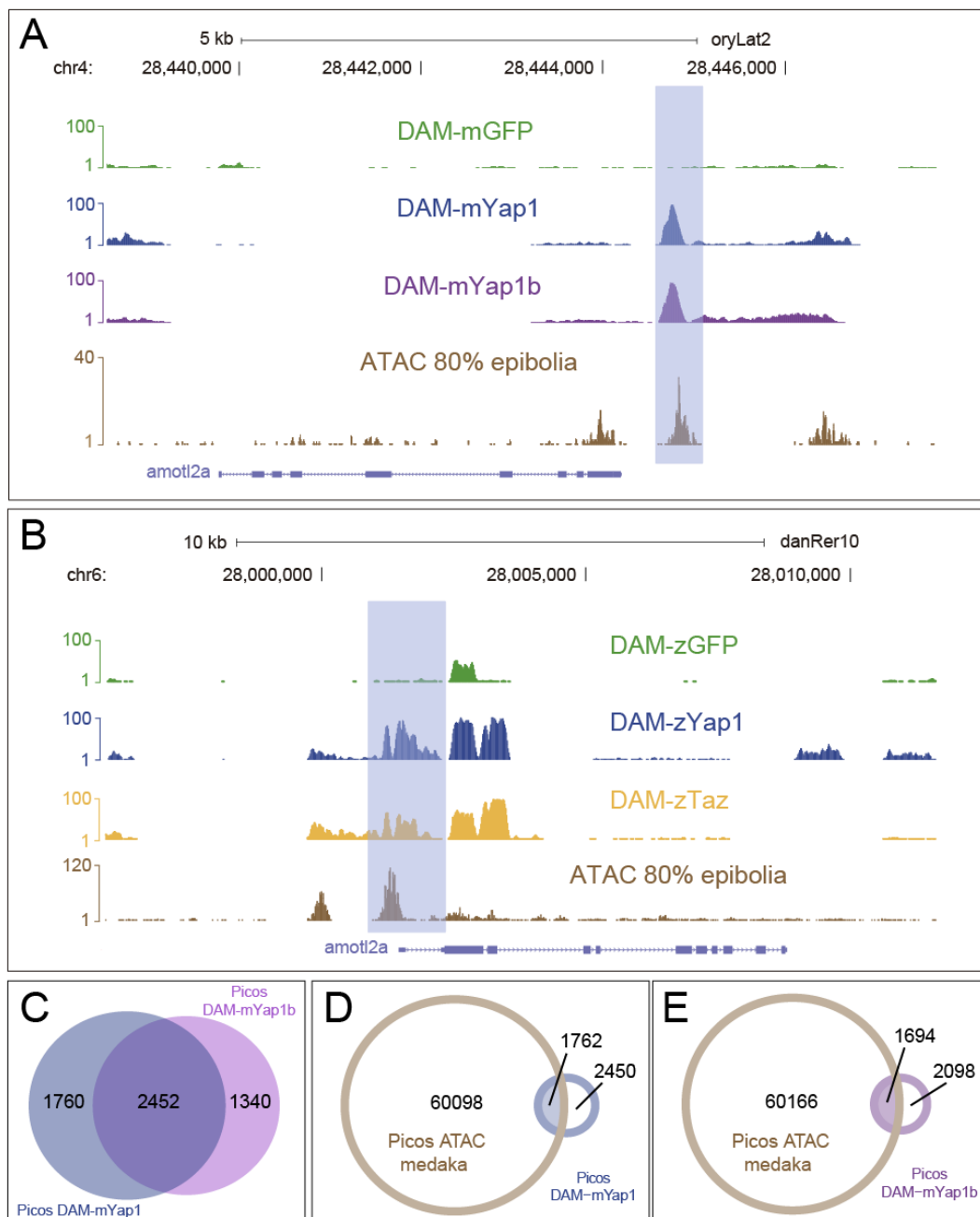


Figura 11 – Perfil de iDamID-seq de proteínas de la familia Yap. Picos de iDamID-seq cercanos al locus del gen *amotl2a*, una diana conocida de Yap, en medaka y pez cebra (**A**, **B**). Se compara el perfil del conjunto de picos DAM-*yap1*, DAM-*yap1b* y DAM-GFP de embriones de medaka en estadio de 80% epibolia (**A**), así como el del conjunto de picos DAM-*yap1*, DAM-*taz* y DAM-GFP de pez cebra en la misma etapa de desarrollo (**B**) con datos de ATAC-seq disponibles para ambas especies. Nótese la presencia de regiones metiladas (áreas sombreadas) en las cercanías del promotor del gen *amotl2a*. Diagramas de Venn que muestran un destacable solapamiento entre picos DAM-*yap1* y DAM-*yap1b* (**C**). También se muestra el nivel de solapamiento entre marcas de ATAC-seq y picos DAM-*yap1* (**D**) y DAM-*yap1b* (**E**).

También se calculó el grado de solapamiento entre los diferentes sets de picos de DamID y marcas de ATAC-seq, disponibles en pez cebra y medaka para el estadio de 80% epibolia, el mismo que el de los embriones utilizados en este análisis (Marlétaz et al., 2018; en revisión).

Aproximadamente, el 42% y el 45% de los picos de *yap1* y *yap1b*, respectivamente, solapan con regiones de cromatina abierta en medaka (Figura 11D-E); por otra parte, el 49% y el 25% de los picos de *yap1* y *taz*, respectivamente, coinciden con marcas de ATAC en pez cebra. El análisis comparativo de la distribución de picos de acuerdo con su distancia al sitio de inicio de la transcripción (*transcription start site*, *TSS*) más cercano revela un enriquecimiento de sitios de unión de proteínas de la familia Yap en promotores y enhancers proximales con respecto a los controles DAM:GFP (Figura Suplementaria 9C-F).

	Nº picos identificados por iDEAR	Nº genes asociados a picos
Medaka Yap1	4212	3112
Medaka Yap1b	3792	2882
Zebrafish Yap1	7553	5427
Zebrafish Taz	8608	5230

Tabla 1 – Picos identificados de iDamID-seq y genes asociados. Número total de picos DamID identificados mediante el programa iDEAR y número de genes asociados al conjunto de picos para cada parólogo de la familia Yap en pez cebra y medaka.

Acto seguido, empleamos la herramienta DREME (Bailey, 2011) para identificar motivos específicos de unión de factores de transcripción al DNA que estuvieran significativamente enriquecidos en la colección de picos obtenidos para cada una de las proteínas de la familia Yap examinadas (Figura 12; Figura Suplementaria 10). Entre los motivos más representativos para las proteínas Yap1 y Yap1b de medaka, el más destacado corresponde a las secuencias consenso de los factores de transcripción TEAD: DREME arroja un E-valor de 5.3×10^{-143} para Yap1 y un E-valor de 5.1×10^{-98} para Yap1b (Figura 12A-B). Un análisis más exhaustivo reveló que una porción significativamente grande del conjunto de picos de Yap1 y Yap1b contenían el motivo TEAD, particularmente aquellas regiones con una puntuación más alta tras aplicar el programa iDEAR (Figura 12C). Además de los factores TEAD, también se detectó, aunque con un menor nivel de significación, la existencia de motivos asociados a otros factores de transcripción conocidos, tanto para el conjunto de picos de Yap1 (p. ej.: RREB1, TBX y ZNF143), como para el conjunto de picos de Yap1b (p. ej.: GATA2, OTX2 y TP73), o bien para ambos (p. ej.: POU y SOX) (Figura 12A-B).

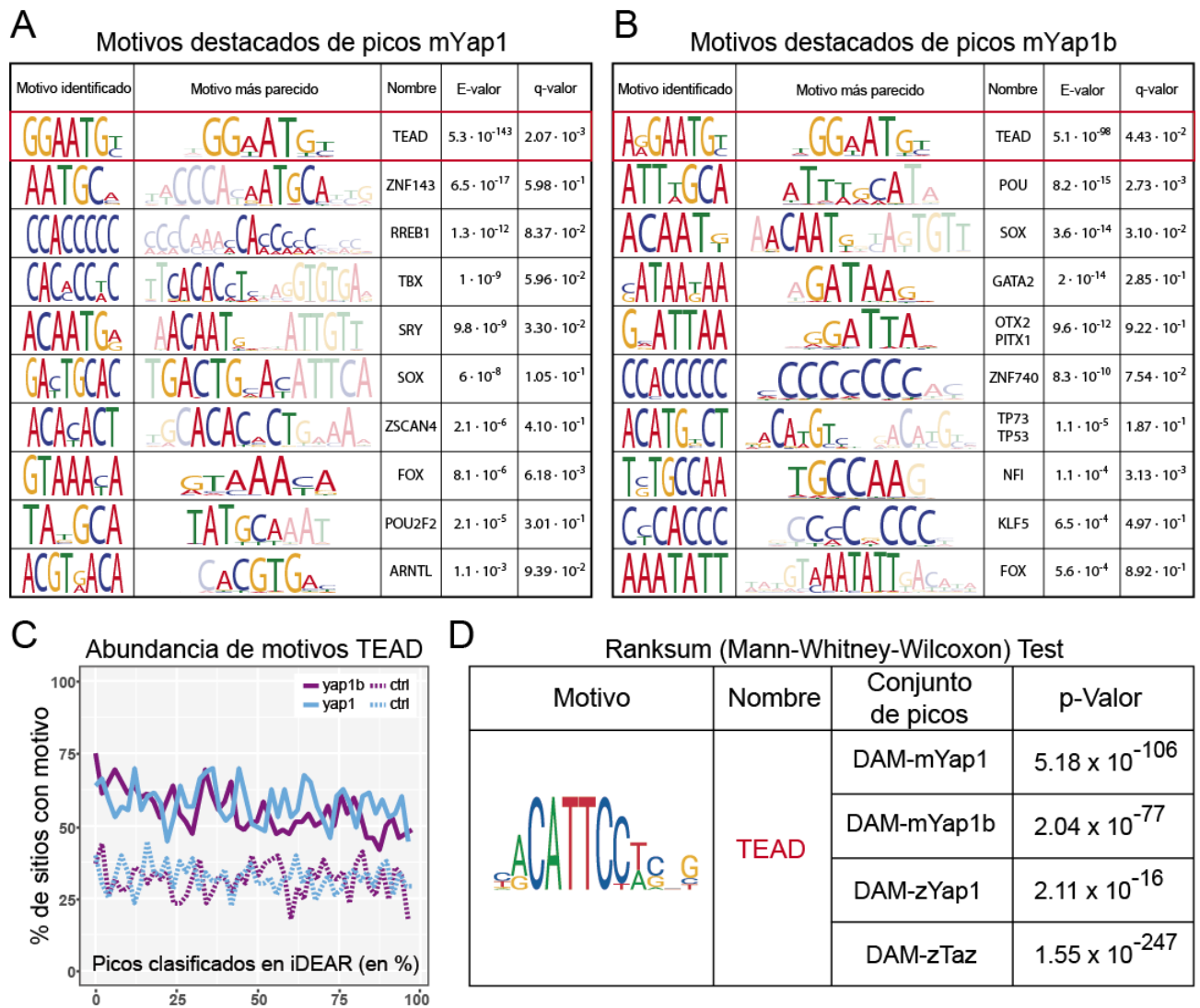


Figura 12 – Identificación de motivos asociados a picos de DamID-seq en medaka. Lista de motivos significativamente asociados a picos DAM-yap1 (A) y DAM-yap1b (B) en medaka identificados por DREME y reconocidos utilizando la herramienta TomTom. Se muestran los E-valores de DREME y los q-valores de TomTom. El motivo TEAD está resaltado en rojo. La abundancia de motivos TEAD en relación a la puntuación otorgada por iDEAR para cada pico se muestra tanto para el conjunto de picos positivos como para el de picos control correspondiente a cada proteína (C). El análisis llevado a cabo por la herramienta AME muestra un significativo enriquecimiento del motivo TEAD en todos los conjuntos de datos DamID-seq identificados (D).

Al contrario que el caso de los parálogos de medaka, la búsqueda de motivos enriquecidos en pez cebra mediante DREME arrojó resultados diferentes al comparar los conjuntos de picos de Yap1 con los de Taz. Los picos de Yap1 de pez cebra muestran enriquecimiento en una serie de motivos conocidos, similares a los obtenidos para ambos parálogos de medaka, incluyendo TEAD (con el segundo grado de significación más alto, un E-valor de 7.4×10^{-147}) y otros como FOX, SOX, TP63 y POU (Figura Suplementaria 10A). Por el contrario, la colección obtenida de motivos asociados a los picos de Taz en pez cebra parece más divergente al resto, con el motivo TEAD en el puesto 22 en cuanto a nivel de significación (E-valor de DREME de 9.5×10^{-3})

(Figura Suplementaria 10B). Esta observación quizá podría sugerir un comportamiento cis-regulador diferente para la proteína Taz en pez cebra, o bien, a una detección incompleta de motivos TEAD por el programa DREME, debido a pequeñas diferencias en la secuencia consenso. Para analizar esta última posibilidad, volvimos a escanear el conjunto de picos de las diferentes proteínas de la familia Yap estableciendo una secuencia TEAD consenso empleando la herramienta AME. Mediante este método, es posible detectar un enriquecimiento muy significativo del motivo TEAD en el conjunto de picos asociados a todos los miembros de la familia Yap, incluyendo a la proteína Taz de pez cebra (Figura 12D).

Con el fin de estudiar la colección de genes asociada potencialmente a las proteínas de la familia YAP, asociamos a cada pico su TSS más cercano (Tabla Suplementaria 3). Al igual que sucedía al comparar el solapamiento entre picos de DamID, al comparar las listas de genes entre parálogos pudimos observar que una importante fracción de genes asociados a Yap1b (aproximadamente el 71%) también se asocian a Yap1 en medaka. Sin embargo, la proporción de genes potencialmente co-regulados por Yap1 y Taz resultó ser más baja (43%) en pez cebra (Tabla 1; Figura Suplementaria 9G-H). Esta diferencia coincide con el alto grado de conservación del dominio de unión a Tead entre Yap1 y Yap1b, que es inferior al existente para dicho dominio entre Yap y Taz (Figura Suplementaria 5), e indica que, a pesar del divergente dominio de transactivación de Yap1b, ambos parálogos de medaka se unen en gran medida a las regiones cis-reguladoras de los mismos genes. Para profundizar en la función de estas proteínas durante el desarrollo, estudiamos las propiedades de los genes potencialmente regulados por las mismas. Para ello, aplicamos la herramienta PANTHER (Mi et al., 2017), con la intención de comprobar la existencia de términos de ontología genética (GO) estadísticamente sobrerrepresentados. Al examinar los términos GO para las categorías de *Función molecular (Molecular function)* y *Tipo de proteína (Protein class)*, se obtuvo una lista de términos representativos similar para aquellos genes asociados a Yap1, a Yap1b o a ambos, incluyendo términos como Actividad de receptores-quinasas de tirosina transmembrana (*Transmembrane receptor tyrosine kinase activity*) y Actividad de factores de transcripción (*Transcription factor activity*) como los más significativos (Tabla 2; Tabla Suplementaria 4). La categoría de términos de GO relacionados con procesos biológicos (*Biological process*) incluye términos de interés como: *Represión de la apoptosis (Negative regulation of apoptotic process)*, *Desarrollo embrionario (Embryo development)*, *Diferenciación celular (Cell differentiation)* y *Regulación del ciclo celular (Regulation of cell cycle)*. Estos términos podrían relacionarse a alguna de las diferentes funciones que se han asociado a miembros de la familia Yap, tales como diferenciación, proliferación y supervivencia celular (Piccolo et al., 2014).

PANTHER GO-Slim Biological Process	Fold Enrichment	raw P value	FDR
negative regulation of apoptotic process	2.66	6.35E-06	7.59E-04
transmembrane receptor protein serine/threonine kinase signalling pathway	2.58	1.05E-03	2.79E-02
embryo development	2.28	6.34E-04	2.16E-02
ectoderm development	2.02	3.64E-05	2.17E-03
transmembrane receptor protein tyrosine kinase signaling pathway	1.95	1.34E-03	3.21E-02
regulation of cell cycle	1.86	2.09E-03	4.55E-02
MAPK cascade	1.76	1.84E-04	7.34E-03
cell differentiation	1.74	6.73E-06	5.36E-04
regulation of phosphate metabolic process	1.59	1.30E-04	6.23E-03
PANTHER GO-Slim Molecular Function	Fold Enrichment	raw P value	FDR
sequence-specific DNA binding transcription factor activity	1.36	5.26E-04	4.92E-02
PANTHER Protein Class	Fold Enrichment	raw P value	FDR
transcription factor	1.52	7.52E-07	1.62E-04

Tabla 2 – Términos GO-Slim identificados de genes asociados a Yap1 en medaka. Esta tabla muestra el nivel de significación estadística (Fold enrichment, p-valor y FDR) para las categorías “Proceso biológico” (*Biological Process*), “Función molecular” (*Molecular Function*) y “Tipo de proteína” (*Protein Class*). El grado de significación se calculó empleando un test de Fisher con un test de corrección múltiple de FDRs.

4. La mutación del gen *yap1b* potencia los defectos en el desarrollo de embriones *yap1*^{-/-} durante la embriogénesis de medaka

Nuestros análisis de iDamID-seq indican que el parólogo *yap1b* controla la expresión de una colección de genes similar a la controlada por *yap1*, a través de numerosos elementos cis-reguladores que coinciden en ambos casos. Este hecho podría sugerir que, a pesar de su divergencia en el dominio de transactivación C-terminal, *yap1b* podría desempeñar un papel complementario a *yap1* en procesos de regulación transcripcional. Como hipótesis alternativa, si el dominio de transactivación atípico de *yap1b* fuera capaz de reclutar al complejo de la Poll 2, la proteína podría actuar como una forma dominante negativa de *yap1*. Para determinar qué explicación resulta más plausible y entender mejor la función de *yap1b*, se generaron mutaciones específicas sobre el exón 2 de este gen en medaka utilizando la tecnología CRISPR-Cas9. Se identificaron dos alelos diferentes (*yap1b*Δ136pb y *yap1b*Δ84pb) y se mantuvieron como líneas estables. Ambas consisten en sendas lesiones cromosómicas de 136 y 84 pb,

respectivamente (Figura 13A). La primera, la del alelo *yap1bΔ136pb*, altera la fase abierta de lectura de la proteína y genera un codón de stop prematuro en el dominio de unión a Tead de la proteína; la segunda, la del alelo *yap1bΔ84pb*, en una delección de 28 aminoácidos en dicho dominio (Figura 13B).

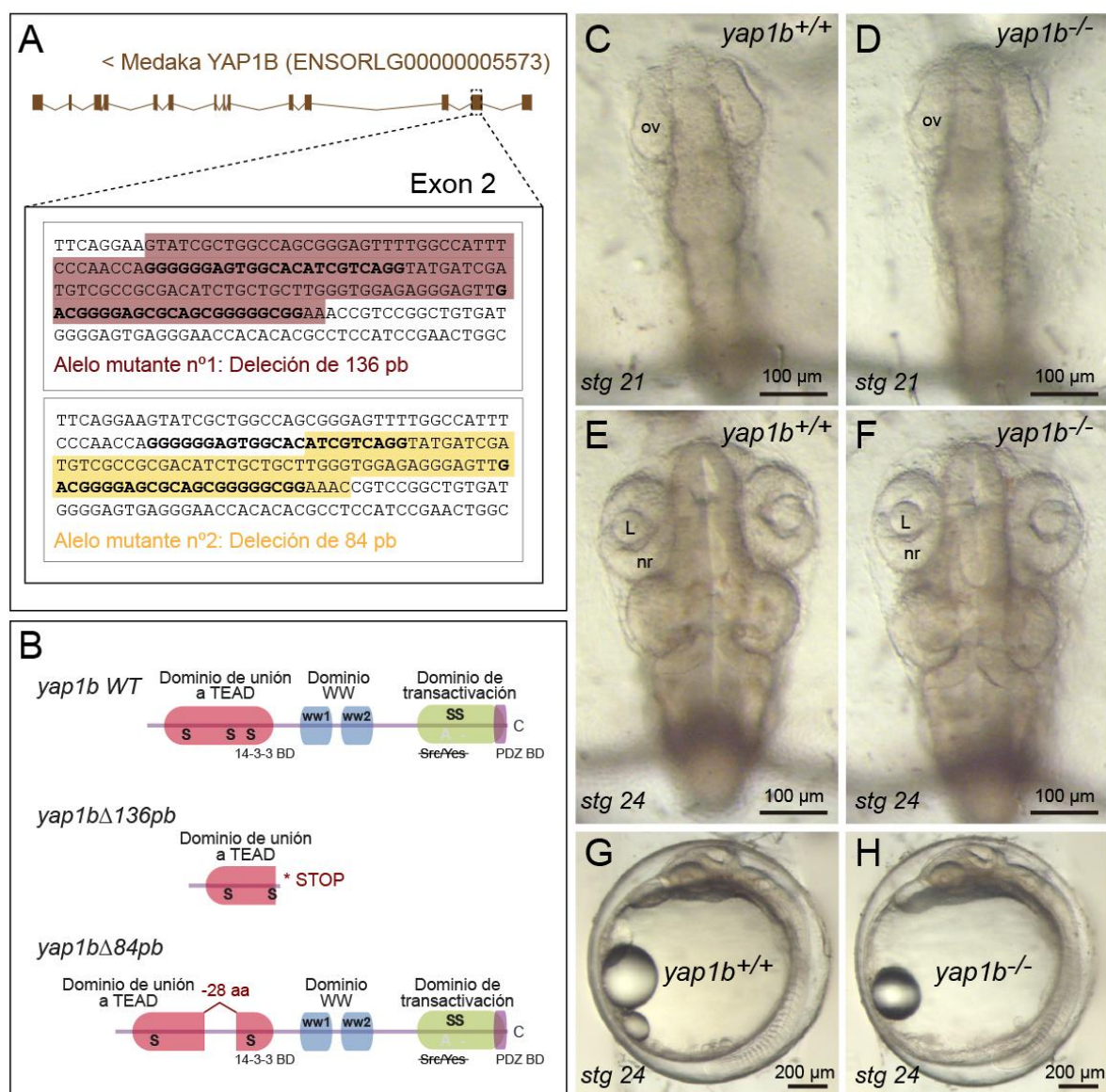


Figura 13 – Generación de alelos mutantes *yap1b* mediante tecnología CRISPR-Cas9. **A)** Representación esquemática de dos delecciones de 136 y 84 pb, generadas en el exón 2 de este gen en medaka. Las secuencias en negrita corresponden a sitios diana de sendos sgRNA (vea *Materiales y métodos*). **B)** Estructura proteica predicha para el gen *yap1b* silvestre y para los dos alelos mutantes obtenidos. La delección *yap1bΔ136pb* genera un cambio en la pauta abierta de lectura y un codón de STOP prematuro, mientras que la delección *yap1bΔ84pb* produce la pérdida de 28 aminoácidos en el dominio de unión al factor de transcripción TEAD. No se aprecian diferencias fenotípicas destacables al comparar embriones WT (**C**, **E**, **G**) con embriones *yap1b*^{-/-} (**D**, **F**, **H**) en las etapas de desarrollo 21 (**C-D**) y 24 (**E-H**). Leyenda: ov: vesícula óptica; L: lente; nr: retina neural.

A pesar de que ambos alelos generan proteínas truncadas o con alteraciones importantes en sus dominios, los embriones *yap1b*^{-/-} se desarrollan con total normalidad y resulta imposible distinguirlos de sus hermanos no mutantes (Figura 13C-H). De hecho, ni la tasa de supervivencia de estos peces en la vida adulta ni su fertilidad se ven comprometidas. Este resultado excluiría la posibilidad de que Yap1b desempeñase un papel principal y no redundante durante la embriogénesis y el mantenimiento de los tejidos. Con el objeto de determinar si este gen juega un papel cooperativo con *yap1*, cruzamos ejemplares fértiles *yap1*^{+/-} *yap1b*^{+/-}, dobles heterocigotos, y examinamos tanto el fenotipo como el genotipo de su progenie. Las observaciones llevadas a cabo nos permitieron llegar a la conclusión de que ambos genes mantienen una interacción epistásica negativa. De esta manera, las malformaciones fenotípicas observadas en mutantes *yap1*^{-/-} (ausencia total de plegamiento de la copa óptica, desalineación de la lente, retraso en el cierre del blastoporo e incremento del grado de muerte celular) se agravan en embriones *yap1*^{-/-} *yap1b*^{+/-} y aún más en mutantes dobles (*yap1*^{-/-} *yap1b*^{-/-}) (Figura 14A-D). El análisis anatómico de estos dobles mutantes bajo el campo claro de un estereoscopio reveló signos de incremento en la tasa de muerte celular del estadio 20-21 en adelante (Figura 14D). Para confirmar este hallazgo, se analizaron los niveles de caspasa-3 en embriones de estadio 20 obtenidos del cruce entre ejemplares *yap1*^{+/-} *yap1b*^{+/-} bajo estereoscopio fluorescente (Figura 14E-H) y microscopio confocal (Figura 14I-L). Estos experimentos de inmunofluorescencia confirmaron un progresivo y significativo incremento en el número de células apoptóticas en embriones *yap1*^{-/-} a medida que se van mutando copias adicionales de *yap1b* (Figura 14Q). Por el contrario, se observaron niveles similares de apoptosis en embriones con al menos una copia viable de *yap1* (Figura Suplementaria 11). De hecho, no se aprecian malformaciones morfológicas en estos embriones, lo que sugiere que una copia de *yap1* es suficiente para rescatar el fenotipo por completo. Por último, se examinó el impacto de la mutación doble en ambos genes en la tasa mitótica de estos embriones, mediante experimentos de inmunofluorescencia específicos contra pH3 en embriones de estadio 20. A pesar del conocido papel desempeñado por las proteínas de la familia Yap en el control del crecimiento de los tejidos y en la proliferación celular (Huang et al., 2005), no fue posible detectar diferencias significativas en el número de células mitóticas por área entre los diferentes genotipos examinados (Figura 14R). Esto sugiere que, al menos en el contexto de embriones en las primeras etapas del desarrollo, los diferentes parálogos de la familia Yap no cumplen ninguna función reseñable en el control de la proliferación celular.

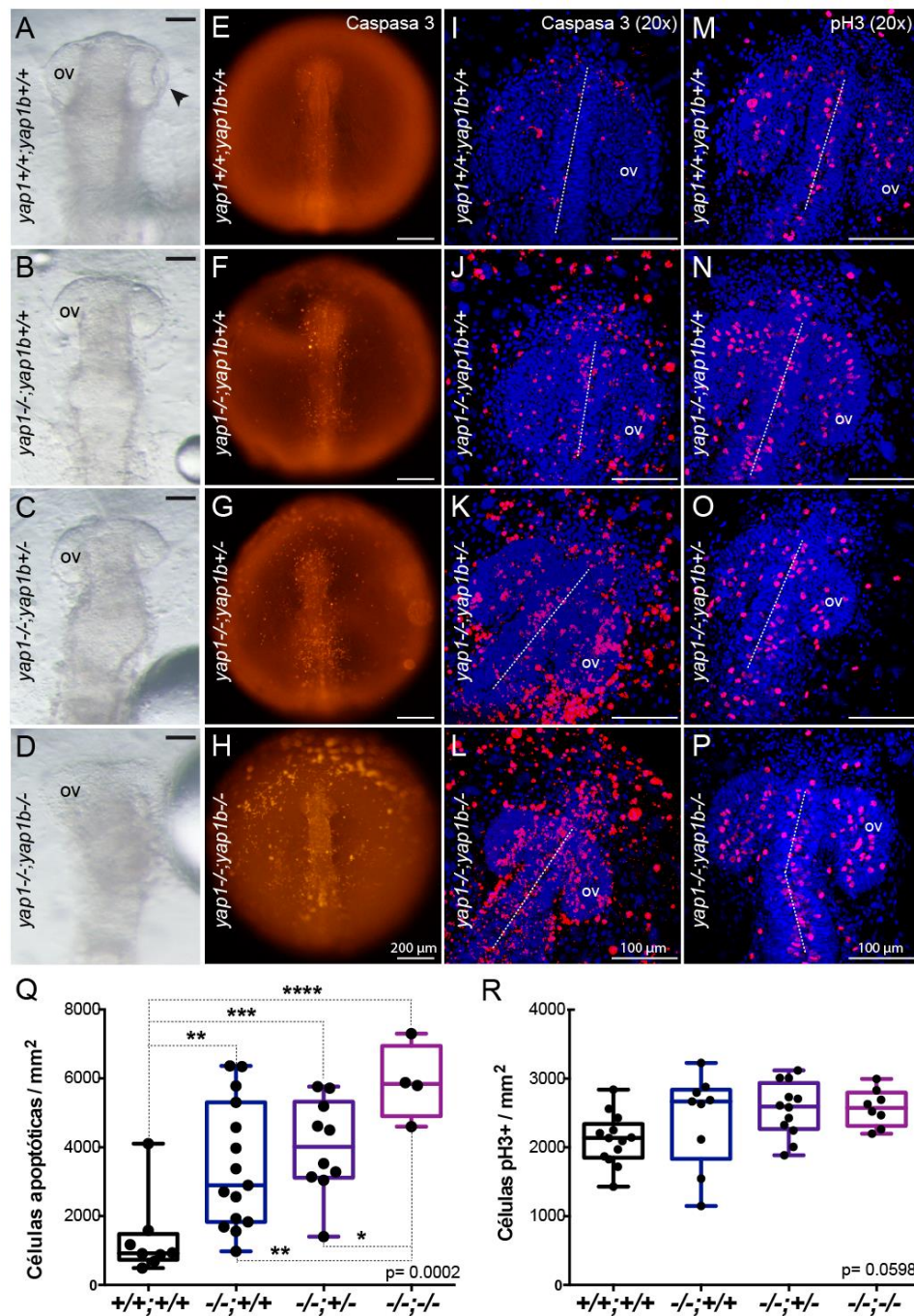


Figura 14 – Niveles de apoptosis y proliferación celular en embriones mutantes *yap1^{-/-} yap1b^{-/-}*. Imágenes de campo claro de la mitad anterior de embriones *yap1^{+/+} yap1b^{+/+}* (A), *yap1^{-/-} yap1b^{+/+}* (B), *yap1^{-/-} yap1b^{+/-}* (C) y *yap1^{-/-} yap1b^{-/-}* (D) en estadio 21, resultado de un cruce de dobles heterocigotos, que revelan una interacción epistática negativa entre *yap1* y *yap1b*. Nótese que las malformaciones observadas en embriones *yap1^{-/-}* son más intensas en embriones *yap1^{-/-} yap1b^{-/-}* (D). Experimentos de inmunofluorescencia con caspasa-3 que revelan la presencia de células apoptóticas en embriones *yap1^{+/+} yap1b^{+/+}* (E, I), *yap1^{-/-} yap1b^{+/+}* (F, J), *yap1^{-/-} yap1b^{+/-}* (G, K) y *yap1^{-/-} yap1b^{-/-}* (H, L) de estadio 20. Se muestran imágenes tomadas de estos embriones en un estereoscopio fluorescente (E-H) y en un microscopio confocal, con tinción de DAPI (I-L). Imágenes de microscopio confocal con tinción de DAPI de células mitóticas pH3 positivas en embriones *yap1^{+/+} yap1b^{+/+}* (M), *yap1^{-/-} yap1b^{+/+}* (N), *yap1^{-/-} yap1b^{+/-}* (O) y *yap1^{-/-} yap1b^{-/-}* (P) en estadio 20. Cuantificación de células positivas en caspasa-3 por área que muestran un significativo aumento en el nivel de apoptosis en embriones *yap1^{-/-}* a medida que se van perdiendo copias del gen *yap1b* (Q). Cuantificación de señales pH3 positivas por área que no muestran diferencias significativas en el nivel de proliferación celular entre genotipos (R). Se realizó un análisis ANOVA (One-way ANOVA) seguido de un test de Fisher entre las diferentes condiciones para evaluar el grado de significación estadística. Leyenda: ov: vesícula óptica.

Por otra parte, la mutación simultánea de ambos genes parece afectar seriamente al proceso de convergencia-extensión de estos embriones en mayor medida que una mutación producida exclusivamente en el gen *yap1* (Figura Suplementaria 12). Se inyectaron embriones del cruce entre heterocigotos *yap1*^{+/-} *yap1b*^{+/-} con un marcador de membrana (LynTdTomato) y se tomaron fotos de estos embriones en el estadio 20 de desarrollo. Se descubrió que todos los embriones mutantes *yap1*^{-/-} *yap1b*^{-/-} presentaban una visible descondensación en su región posterior, también presente en algunos embriones *yap1*^{-/-} *yap1b*^{+/-}. Sin embargo, dicha anomalía no resulta apreciable en mutantes *yap1*^{-/-} ni por supuesto en embriones WT. En resumen, nuestros resultados sugieren que los parálogos *yap1* y *yap1b* cooperan en la regulación de procesos de supervivencia celular y en la morfogénesis de tejidos.

5. Identificación de genes regulados por *yap1* durante la gastrulación tardía de medaka

Para profundizar en la comprensión de la regulación transcripcional coordinada por *yap1* durante las etapas tardías de la gastrulación y así complementar los datos recabados mediante iDamID-seq (mostrados en el apartado 3 de esta sección), se llevó a cabo un estudio identificativo de las dianas transcripcionales de esta proteína en embriones de medaka de estadio 16 mediante RNA-seq (Figura 15). Embriones *yap1*^{-/-} y *yap1*^{+/-} fueron procesados por separado para extraer tanto su fracción de RNA como su DNA genómico. Esta última fracción fue utilizada para genotipar cada embrión y así obtener material suficiente para cada condición. Se enviaron tres réplicas biológicas para cada una de estas condiciones para su secuenciación en plataformas NGS. Las lecturas obtenidas fueron mapeadas contra el genoma de medaka (*oryLat2*) y, tras determinar el número de copias para cada gen en la condición mutante, ésta fue comparada con respecto al número de transcritos de cada gen en el control WT para trazar la relación de genes diferencialmente expresados (reprimidos y activados) en embriones mutantes con respecto a embriones WT. Se identificaron 188 genes subregulados y 105 genes sobreexpresados en los embriones mutantes (Figura 15A; Tabla Suplementaria 5). La lista de genes obtenida con un menor nivel de expresión en el mutante es particularmente interesante. Entre ellos, destacamos genes pertenecientes a la propia ruta Hippo y a otros mecanismos de regulación (p. ej.: *yap1*, *lats2*, *src*, *amotl2a*, *amotl2b* y *TEAD4*), genes que codifican para componentes de la matriz extracelular o que forman parte de complejos de adhesión (p. ej.: *ctgfa*, *cyr61*, *frem1a*, diversos tipos de colágeno y varios tipos de integrinas), genes involucrados en la reorganización del citoesqueleto y en la migración celular (p. ej.:

acta1a, *marcksl1b*, *myo9ab* y *ptk2bb*), genes implicados en procesos de apoptosis (p. ej.: *boka*, *PEA15* y *pim3*) o genes que codifican para factores de transcripción o proteínas involucradas en otras rutas de señalización (p. ej.: *cxcl14*, *SHISA2*, *foxo1*, *tp63* y *bmp4*).

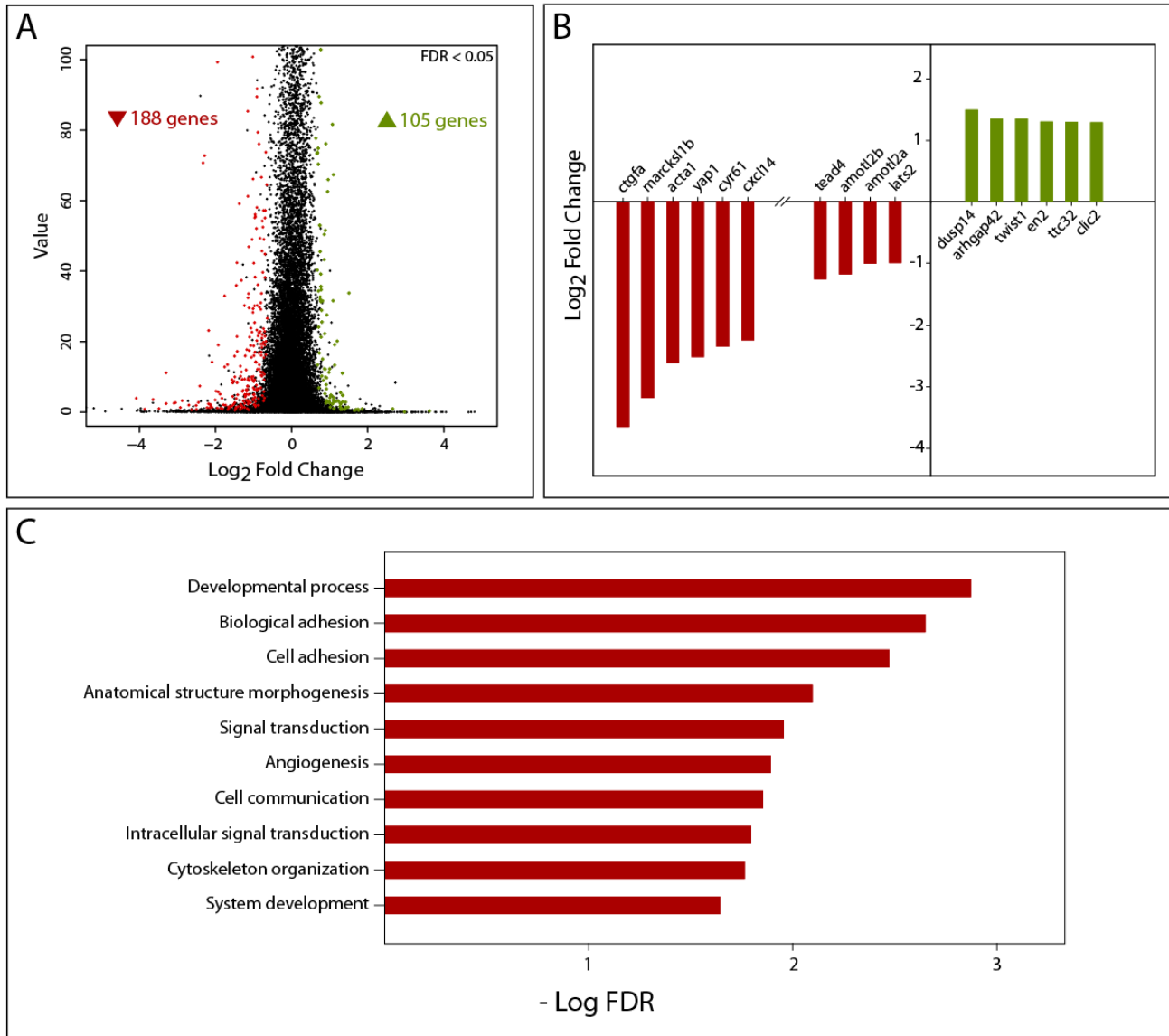


Figura 15 – Identificación de dianas transcripcionales de *yap1* mediante RNA-seq. Un análisis comparativo en el nivel de expresión genético entre embriones *yap1*^{+/+} y embriones *yap1*^{-/-} permite identificar el conjunto de genes controlados por *yap1*. Se identificaron 188 genes desregulados y 105 genes sobreexpresados en los embriones mutantes (**A**). Entre los genes con un menor nivel de expresión en embriones mutantes con respecto a embriones WT destacan algunas dianas identificadas en experimentos de iDamID-seq, como *ctgfa*, *yap1* y *amotl2a*. El gen *ctgfa* es el que experimenta un mayor cambio, pues se expresa a una tasa más de 6 veces menor en embriones mutantes (**B**). Análisis de términos GO asociados a la lista de 188 genes desregulados en embriones mutantes, que revela la implicación de la mayoría de estos genes en procesos de adhesión y migración celular (**C**).

El gen cuya expresión experimenta un mayor cambio es *ctgfa*, cuyo nivel de expresión decae más de 6 veces en embriones mutantes con respecto a embriones WT (Figura 15B). De entre estos 188 genes menos expresados en el mutante, 73 están asociados a picos DamID cercanos, incluyendo algunos muy importantes, como *ctgfa*, *yap1*, *src*, *amotl2a* y *lats2*, lo que supone aproximadamente el 40% del total (Tabla Suplementaria 6). También existe cierto solapamiento entre esta lista de genes y el conjunto de genes asociados a picos DAM-*Yap1b*, aunque éste resulta menor (aproximadamente el 33%). Se llevó a cabo un análisis de términos GO para determinar la función potencial de estos genes controlados por *yap1* siguiendo el mismo procedimiento que en el apartado 3 (Figura 15C). Se identificaron términos de mucho interés como adhesión celular (*cell adhesion*), transducción de señales (*signal transduction*), comunicación celular (*cell communication*) y organización del citoesqueleto (*cytoskeleton organization*). También se analizó el patrón de expresión de algunos de estos genes en pez cebra recabando información de la base de datos ZFIN y se descubrió que muchos de ellos, incluyendo *yap1*, se expresan con especial intensidad en la línea media del embrión (Figura suplementaria 13). En resumen, el elevado solapamiento entre datos de iDamID-seq y RNA-seq y la función llevada a cabo por el conjunto de genes identificados en este análisis indican que Yap1 controla la expresión de genes que codifican para proteínas implicadas en su propia regulación, incluyendo al propio *yap1*, y la de genes implicados en procesos de adhesión y migración celular, de gran importancia durante la gastrulación no sólo en teleósteos sino también en vertebrados en general.

6. Yap1 cumple un papel importante en el establecimiento del eje corporal de medaka durante la gastrulación

Además de las anomalías morfológicas observadas en los embriones mutantes de medaka *yap^{-/-}* mencionadas anteriormente (Figura suplementaria 3), existen otros rasgos fenotípicos que sugieren defectos en la gastrulación de estos embriones (Figura 16), tales como un eje antero-posterior más corto (Figura 16A-B) y un eje lateral más ancho, que evidencia un ensanchamiento de los somitas a lo largo de dicho eje (Figura 16C-D). Estas anomalías también están presentes en embriones mutantes *yap^{-/-} taz^{-/-}* de pez cebra (Figura Suplementaria 14), lo que sugiere que las proteínas de la familia Yap parecen cumplir alguna función como reguladores de la gastrulación de teleósteos en general, y del proceso de convergencia-extensión en particular. Por lo tanto, para ahondar en esta cuestión, se analizó más en detalle el papel de estas proteínas durante el desarrollo temprano mediante diversos experimentos.

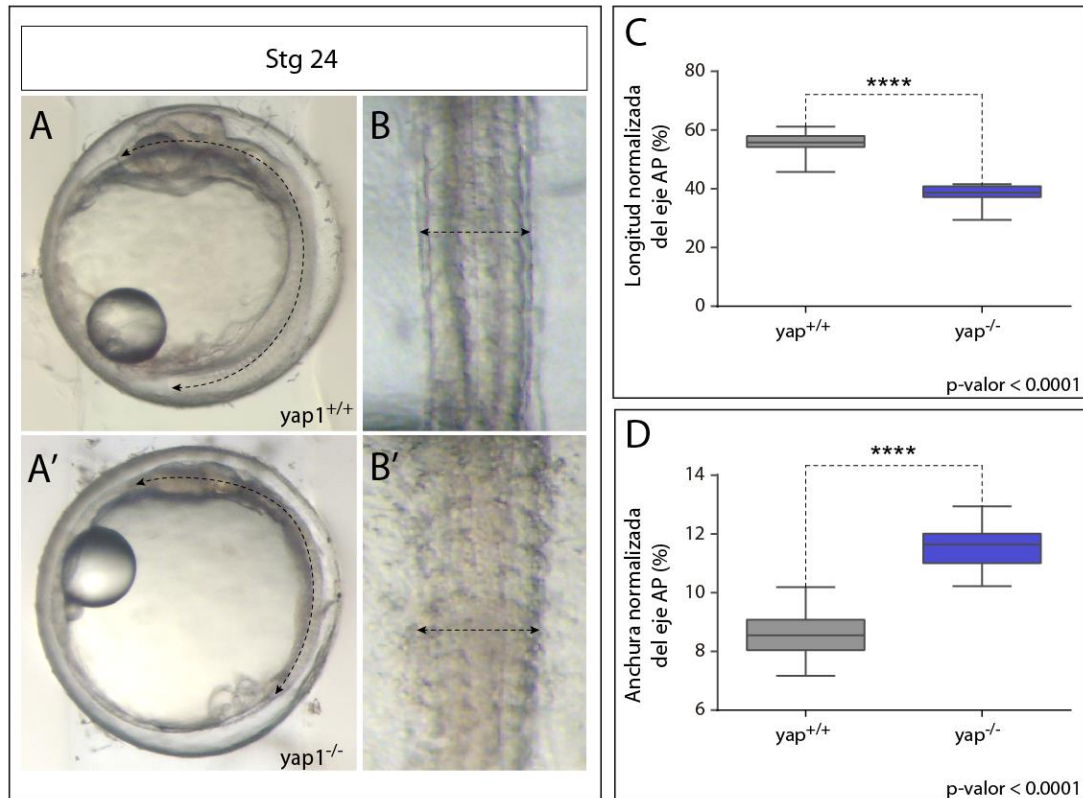


Figura 16 – Longitud y anchura de embriones *yap1*^{-/-} de medaka. Los embriones WT de medaka presentan un eje anteroposterior más largo (**A**) y más estrecho (**B**) que los embriones *yap1*^{-/-} (**A'**, **B'**). La normalización de la longitud del eje AP con respecto a la circunferencia del yolk (**C**) y la del eje lateral con respecto al diámetro del yolk (**D**) certifica cuantitativamente las diferencias fenotípicas entre ambos embriones. Se llevó a cabo un test de Fisher para determinar el grado de significación estadística de estas diferencias anatómicas.

En primer lugar, llevamos a cabo hibridaciones in situ (ISH) de embriones WT y *yap1*^{-/-} de medaka en los estadios 16 (cuando la epibolia ha alcanzado el 80% de la superficie del embrión) (Figura 17) y 19 (una vez que la gastrulación ha finalizado con el cierre del blastoporo y el establecimiento de los principales ejes corporales del embrión) (Figura Suplementaria 15). Como marcador, se utilizaron sondas elaboradas a partir de la secuencia codificante del gen *ntl*, uno de los principales reguladores de la especificación del mesodermo, la formación de la notocorda y el correcto establecimiento de la región caudal en pez cebra (Schulte-Merker et al., 1994), lo que lo hace especialmente apto para detectar posibles anomalías en el proceso de convergencia-extensión en embriones de teleósteos. Los resultados de esta ISH demuestran una evidente deslocalización de las células del mesodermo en embriones *yap1*^{-/-} en ambos estadios, lo que podría sugerir un defecto en la migración de dichos precursores durante la convergencia-extensión de estos embriones.

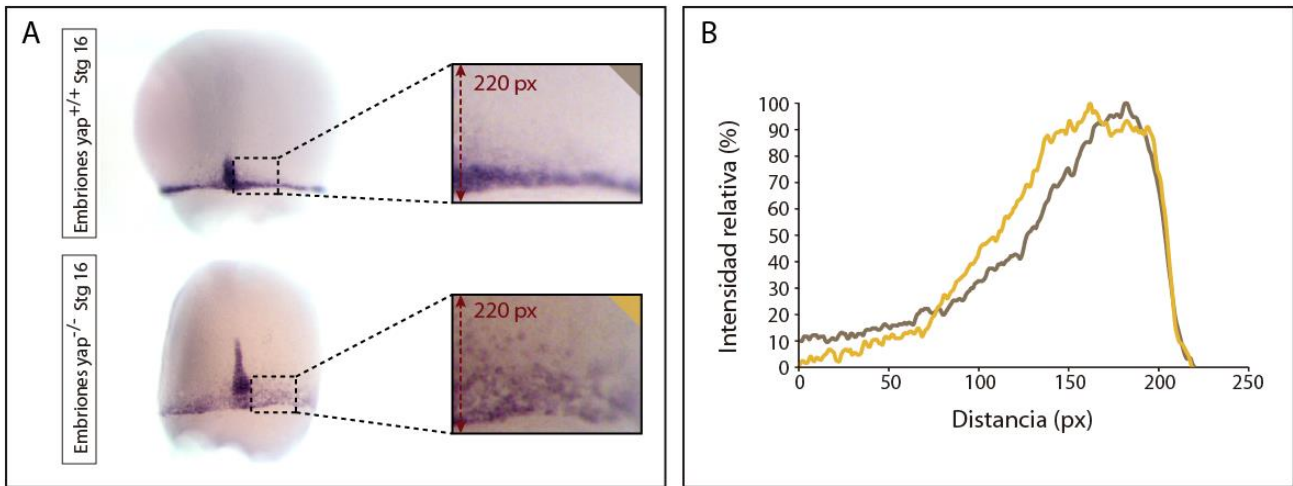


Figura 17 – Patrón de expresión de *ntl* en embriones *yap1*^{+/+} y *yap1*^{-/-} en embriones de medaka de estadio 16. A) El patrón de expresión del gen *ntl*, uno de los especificadores del mesodermo, aparece mucho más definido en embriones *yap1*^{+/+}. B) Gráficas de intensidad de píxel de imágenes ampliadas de los experimentos de hibridación in situ en embriones WT (marrón) y mutante (naranja). Se midió la intensidad de píxel de una sección longitudinal de 220 px de longitud, próxima a la línea media en embriones de ambos genotipos (vea *Materiales y métodos*). El patrón observado en embriones mutantes es más ancho, lo que revela una deslocalización de precursores mesodérmicos.

Para comprender la naturaleza molecular de estas anomalías, quisimos analizar el papel de los genes *ctgfa* y *cyr61* durante la gastrulación de teleósteos. Ambos genes son conocidas dianas transcripcionales de proteínas de la familia Yap (Piccolo et al., 2014), y ya fueron anteriormente identificadas tanto en los experimentos de iDamID-seq como en los de RNA-seq, descritos en apartados previos de este trabajo, como dianas directas en el contexto del desarrollo temprano. Ambas proteínas son parálogas, forman parte de la familia de proteínas extracelulares CCN, para las cuales se han descrito funciones en adhesión celular, migración, proliferación, diferenciación y apoptosis (Malik et al., 2015). Así, un descenso en el nivel de expresión endógeno de *ctgf* en pez cebra provoca deformaciones en la notocorda, retrasos en el desarrollo y un acortamiento de la cola, entre otros defectos (Chiou et al., 2006). Por su parte, *cyr61* participa en procesos tales como el ensamblaje de la matriz extracelular, el establecimiento de adhesiones celulares y la regulación de rutas intracelulares en *Xenopus* (Latinkic et al., 2003). Por esta razón, examinamos el efecto producido por el descenso en los niveles endógenos de ambas proteínas mediante experimentos de *knock-down* empleando morfolinós específicos (Figura 18). La coinyección de morfolinós contra *ctgfa* y *cyr61* provoca un retraso en la epibolia y un acortamiento del eje antero-posterior del embrión hacia el final de la gastrulación (estadios 16-17). Este defecto se hace especialmente patente a medida que el desarrollo avanza (estadio 24).

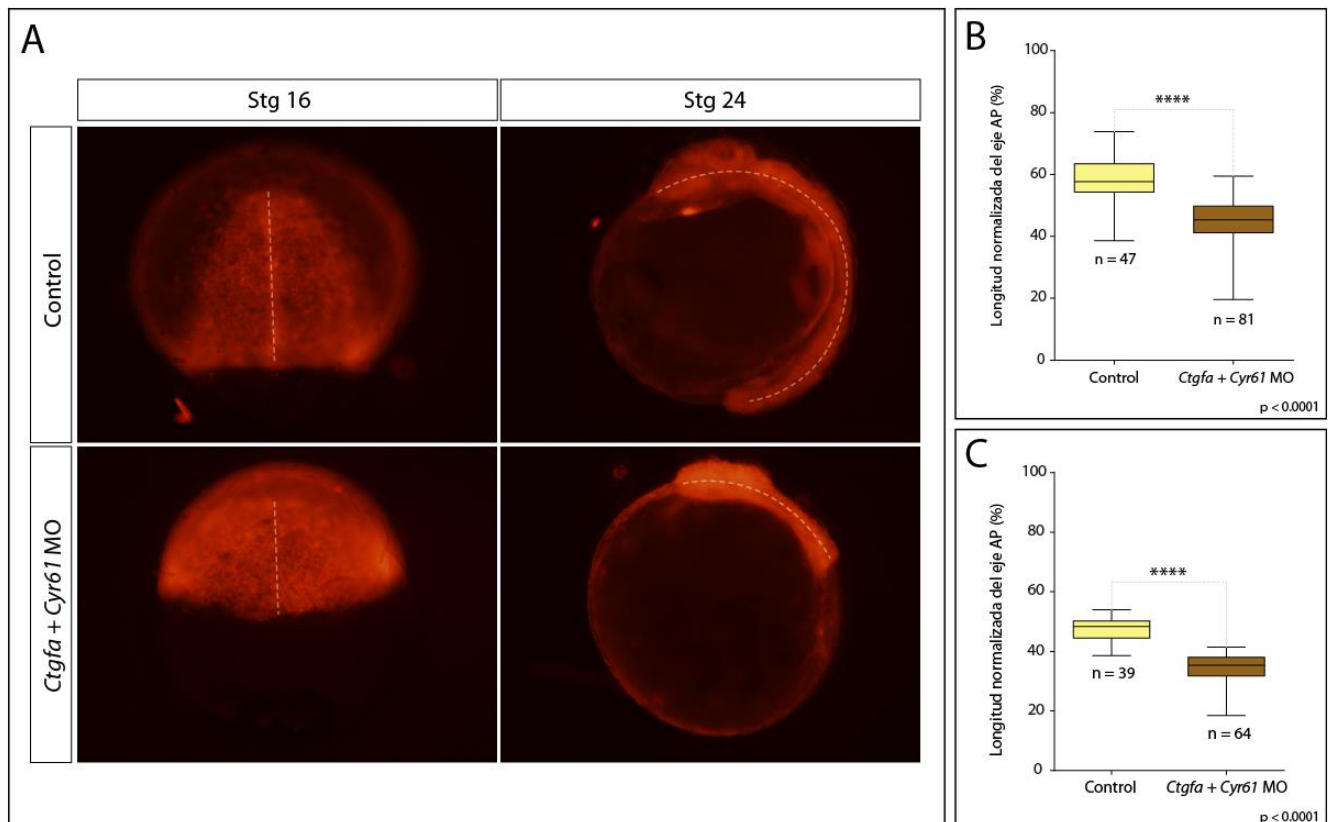


Figura 18 – Inyección conjunta de morfolinolinos contra *ctgfa* y *cyr61* en medaka. La coinyección de morfolinolinos específicos contra los genes *ctgfa* y *cyr61* provoca un severo acortamiento del eje anteroposterior (línea discontinua blanca) en el estadio 16, que se mantiene en las siguientes etapas del desarrollo. Al cabo de 48 hpf, en el estadio 24, el acortamiento de este eje se hace tan patente como la existencia de otras anomalías anatómicas. Ambos morfolinolinos fueron coinyectados junto con mRNA del marcador de membrana LynTdTomato (**A**). Gráficas que demuestran el acortamiento significativo del eje anteroposterior en embriones morfantes en los estadios 16 (normalizado con respecto al diámetro del vitelo) (**B**) y 24 (normalizado con respecto a la circunferencia total del vitelo) (**C**). Se llevó a cabo un test de Fisher para determinar el grado de significación estadística de estas anomalías.

Ambos morfolinolinos fueron inyectados también por separado (Figura Suplementaria 16). Mientras que la disminución del nivel de *ctgfa* parece generar los mismos efectos en el embrión que la coinyección simultánea de ambos morfolinolinos, no parece que la disminución del nivel de *cyr61* por sí sola produzca anomalías destacables en el embrión. No obstante, a medida que avanza el desarrollo, la coinyección de ambos morfolinolinos se traduce en embriones con un eje más corto en comparación con embriones que fueron inyectados únicamente con el morfolino contra *ctgfa* (Figura Suplementaria 16A-C). Por tanto, parece que la expresión de *ctgfa* parece tener un mayor peso en el proceso de gastrulación en embriones de medaka. La amplificación por PCR de las regiones afectadas por estos morfolinolinos permite detectar la adquisición del intrón 1 en el transcrito de *ctgfa* y tanto la adquisición del intrón 2 en algunas

células como la pérdida del exón 2 en otras en el transcrito de *cyr61* (Figura Suplementaria 16D).

Alternativamente, se está tratando de establecer una línea CRISPR sobre ambos genes en peces medaka, hecho que ya se ha conseguido para el gen *ctgfa*. Hasta el momento, se ha obtenido un alelo mutante, consistente en una doble delección de 10 y 13 pb sobre el exón 2 de la proteína, que genera una versión truncada de la misma (Figura Suplementaria 17). Aún no ha sido posible determinar con total exactitud y fiabilidad el fenotipo de estos embriones mutantes. No obstante, la mutación del gen *ctgfa* produce en la generación F0 un elevado porcentaje de embriones CRISPrantes, embriones que acaban desarrollando exactamente las mismas anomalías morfológicas que las observadas en los morfantes (Figura 19). Algunos de estos embriones acaban desarrollando también un segundo eje, pero en un porcentaje muy bajo.

Además del uso de morfolinós, también se optó por llevar a cabo el experimento contrario: sobreexpresar *ctgfa* en embriones de medaka mediante inyección de su mRNA. Con el objeto de crear gradientes asimétricos de esta proteína y así estimar cómo influiría este desequilibrio en la migración de las células durante la gastrulación, las inyecciones se llevaron a cabo en un solo blastómero en estadios de desarrollo de 2, 4 y 8 células (Figura Suplementaria 18). El efecto generado es bastante sorprendente: a las 24 hpf, cuando los embriones se encuentran en estadio 16-17, un elevado porcentaje de los embriones que fueron inyectados con la proteína muestran un eje anteroposterior significativamente más corto, con un frente de epibolia cóncavo, en forma de “V”. Los embriones con un fenotipo más severo logran sobrevivir hasta el estadio 24, pero se aprecia en ellos múltiples anomalías morfológicas, como una expansión del territorio anterior del embrión e incluso el desarrollo de un tercer ojo. La conclusión de estos experimentos es que tanto la pérdida de función de *ctgfa* como su ganancia resultan en defectos morfogenéticos graves durante la gastrulación.

Para comprender mejor el papel del gen *ctgfa* en los complejos movimientos morfogenéticos que caracterizan la gastrulación, examinamos *in vivo* estos movimientos de convergencia-extensión y epibolia mediante microscopía confocal “time-lapse”. Para poder trazar el desplazamiento de las células, se inyectó un marcador nuclear (H2B-GFP) tanto en embriones WT como en embriones morfantes para *ctgfa* y *cyr61*, así como en CRISPrantes para *ctgfa*, embriones F0 inyectados con la tecnología CRISPR para generar mutaciones.

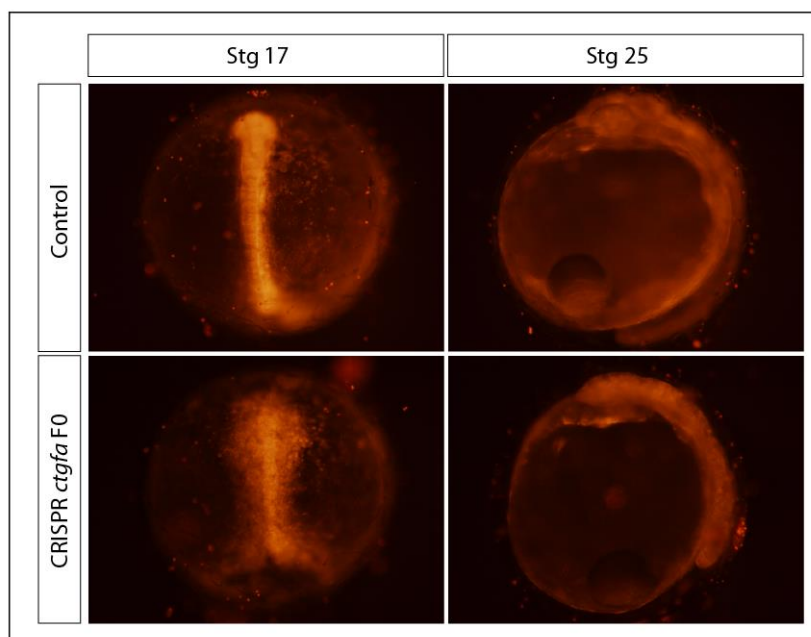
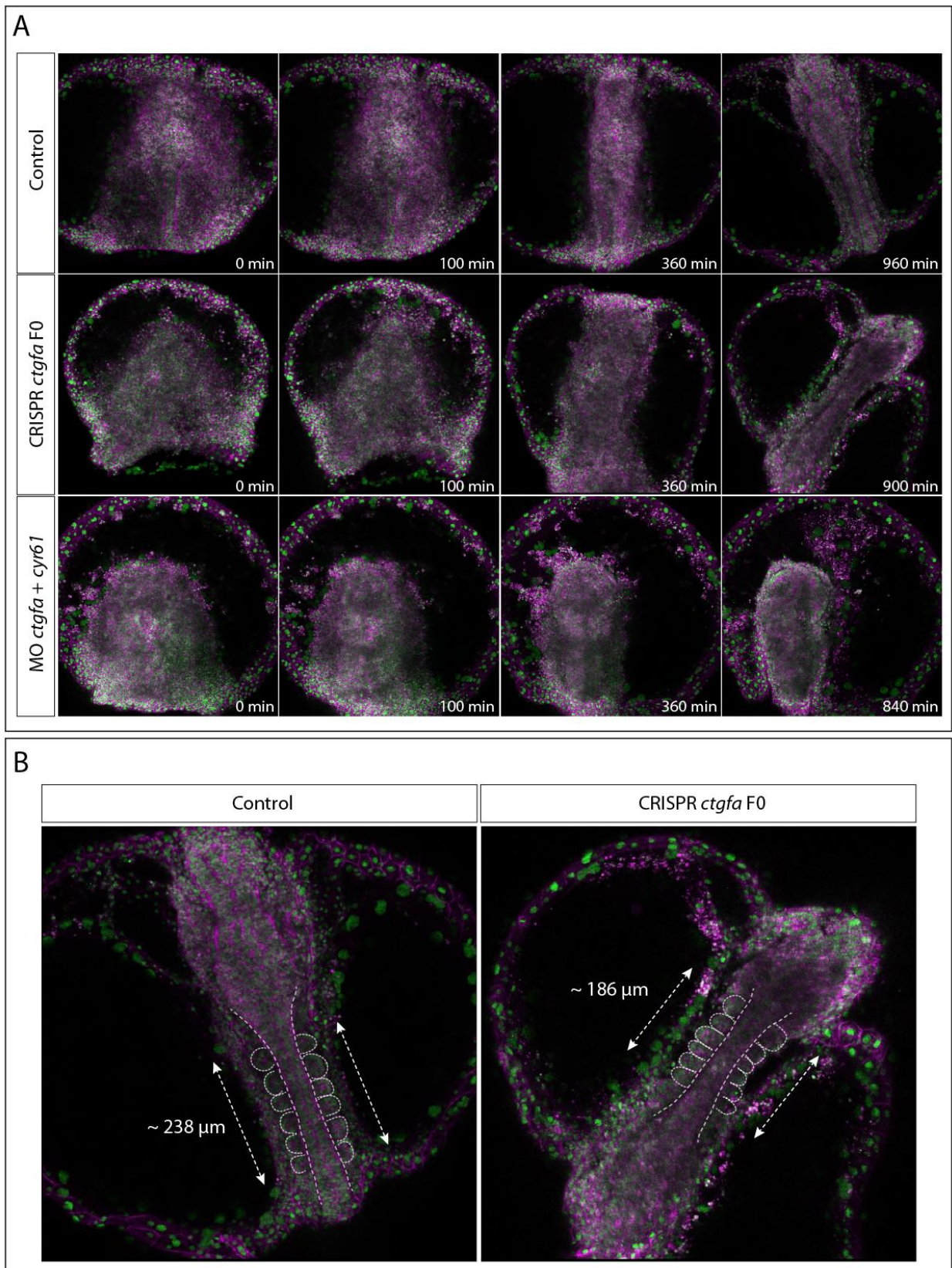


Figura 19 – Obtención de embriones CRISPantes de medaka para el gen *ctgfa*. La mutación del gen *ctgfa* de medaka mediante el uso de la tecnología CRISPR-Cas9 genera embriones F0 con un eje antero-posterior más corto, apreciable desde las etapas finales de la gastrulación, de manera similar a lo que se observa en embriones morfantes para dicho gen.

Así, mientras que en los embriones WT se aprecia una gastrulación completamente normal, donde las células precursoras se intercalan para establecer la línea media del embrión a medida que la epibolia progresa, en los embriones CRISPantes y morfantes el proceso de epibolia es anormal y existe una importante descoordinación entre los procesos de convergencia y extensión, que se producen a un ritmo anormalmente bajo (Figura 20A; Películas Suplementarias 1-3). Estas anomalías se traducen en un acortamiento del eje anteroposterior, tal y como se mencionó anteriormente: la distancia entre el primer y el sexto par de somitas es un 22% más pequeña en embriones CRISPantes con respecto a embriones control (Figura 20B). Para profundizar en el análisis de estas anomalías, se determinó la trayectoria y velocidad de un grupo de células para embriones de las tres condiciones experimentales, tanto en la parte posterior del embrión, cerca del frente de epibolia, como en su parte cefálica. En la parte posterior de los embriones control se observa que las células tienden a desplazarse a la línea media al mismo ritmo que prosigue su migración hacia el polo vegetal. Por el contrario, en los embriones CRISPantes, las células examinadas se desplazan hacia el polo vegetal, pero prácticamente no muestran ningún tipo de desplazamiento hacia la línea media, evidenciando problemas de convergencia. Las células de los embriones morfantes parecen mostrar cierta convergencia hacia la línea media, pero no parece en general tan

intensa como en los embriones control, y la tasa de migración hacia el polo vegetal es prácticamente nula (Figura 21A; Figura suplementaria 19; Películas Suplementarias 4-6).



Para determinar el grado de migración direccional de los precursores examinados, se calcularon dos parámetros: el cociente o ratio de direccionalidad y el desplazamiento cuadrático medio (MSD). En el caso que nos ocupa, el ratio de direccionalidad de las células de los embriones control es más elevado que el de las células de las otras dos condiciones experimentales (Figura 21B). Por otra parte, las células de los embriones control presentan un valor MSD que duplica al de otras condiciones experimentales al cabo de 60 minutos (MSD de $1032 \mu\text{m}^2$ en embriones control por un MSD de $492 \mu\text{m}^2$ en embriones CRISPantes y uno de $411 \mu\text{m}^2$ en embriones morfantes) (Figura 21C). En cuanto a la velocidad media de los precursores analizados, no existen diferencias significativamente palpables entre las tres condiciones experimentales (Figura 21D). Por tanto, al suprimir la función de las proteínas CCN, en especial de *Ctgfa*, la migración direccional de las células en gastrulación hacia la línea media se ve muy comprometida.

En la región anterior del embrión, la situación observada no es muy diferente. En embriones control ya no se aprecia prácticamente ningún tipo de desplazamiento hacia el polo vegetal, pero sí cierta convergencia hacia la línea media, que en cualquier caso es mucho menos intensa que en la región posterior, en consonancia con lo que se deduce al observar las películas que muestran la gastrulación completa del embrión. Sin embargo, en embriones CRISPantes, este mínimo movimiento de convergencia casi no existe. Por el contrario, las células de los embriones morfantes parecen seguir una trayectoria similar a las de las células control, aunque algo más acusada (Figura Suplementaria 20; Figura Suplementaria 21A; Películas Suplementarias 7-9). Su cociente de direccionalidad y su MSD también son ligeramente superiores al de las otras dos condiciones, así como su velocidad media (Figura Suplementaria 20B-D). Aun así, estos valores de persistencia siguen siendo muy bajos en comparación con los de las células control observadas en la región posterior del embrión, en donde el proceso de convergencia-extensión todavía no ha alcanzado su cénit.

(Página anterior) Figura 20 – Películas generales de gastrulación en medaka. A) Time lapse de embriones control, CRISPantes y morfantes de *ctgfa* y *cyr61* en estadio de 90% epibolia durante un periodo de tiempo máximo de 16 horas. Obsérvese que, 6 horas después del inicio de cada time lapse (360 minutos), los embriones control son claramente más estrechos que los embriones CRISPantes y morfantes, lo que sugiere un retraso en la convergencia-extensión o bien una convergencia-extensión incompleta. **B)** Las anomalías en el proceso de convergencia-extensión se ponen de manifiesto en los CRISPantes hacia el final del time-lapse. La distancia comprendida entre el primer y el sexto par de somitas es un 22% inferior en el embrión CRISPante que en el embrión control. Todas las películas fueron tomadas con el mismo nivel de zoom. Señal fluorescente verde: H2B-GFP; Señal fluorescente rosa: LynTdTomato.

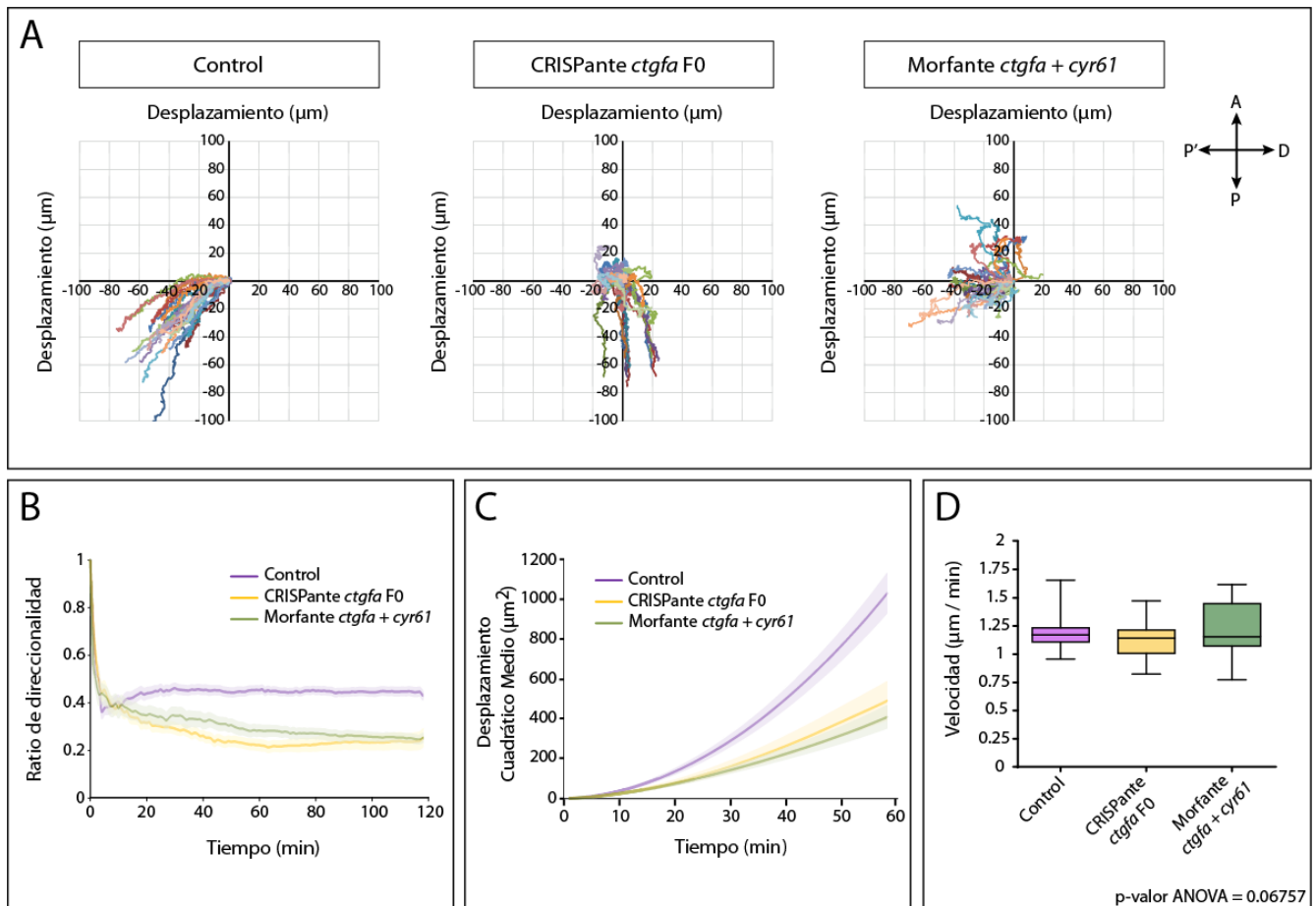
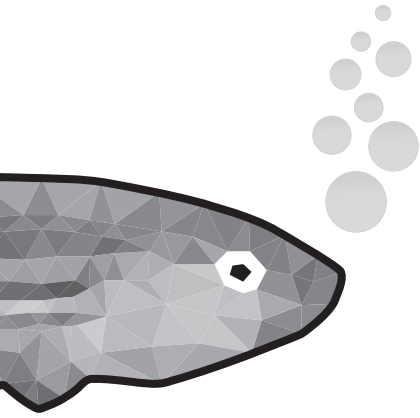
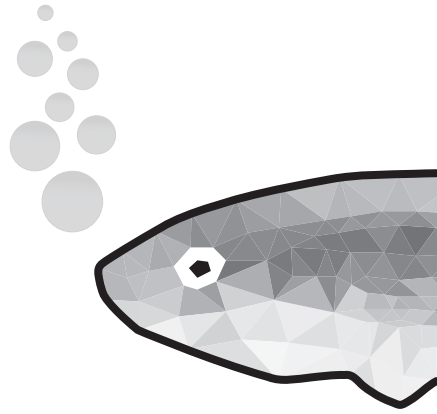


Figura 21 – Rastreo de trayectorias celulares en la región posterior de embriones de medaka durante la gastrulación. A) Conjunto de trayectorias de una muestra de 30 células correspondientes a tres embriones WT, tres CRISPs de *ctgfa* y tres morfantes de *ctgfa* y *cyr61* de estadio 16 a lo largo de un lapso de tiempo de 2 horas, tomadas de películas grabadas in vivo en un microscopio confocal. Mientras que todas las células WT tienden a desplazarse al polo vegetal al mismo ritmo que convergen hacia la región más proximal de la línea media, las de los embriones CRISPs y las de los morfantes muestran trayectorias erráticas. Las células de los embriones CRISPs prácticamente no convergen hacia la línea media, mientras que las de los morfantes casi no se desplazan a la parte posterior del embrión. **B)** Ratio de direccionalidad de células control (púrpura), CRISPs (naranja) y morfantes (verde). El ratio superior de las células control indica que estas células se desplazan en línea recta con una tendencia superior al de las otras dos condiciones. **C)** Desplazamiento cuadrático medio (MSD) de células control, CRISPs y morfantes durante los primeros 60 minutos. Las células control presentan un MSD muy superior. **D)** Velocidad media de células control, CRISPs y morfantes. La velocidad no varía significativamente entre condiciones. En las gráficas B y C, el trazo grueso de cada condición determina el valor promedio, mientras que las regiones sombreadas adyacentes representan su error estándar. Para el análisis de velocidad celular, se llevó a cabo un análisis One-way ANOVA para determinar el grado de significación de las diferencias observadas de velocidad entre las tres condiciones. Leyenda: A: región anterior; P: región posterior; P': extremo proximal a la línea media; D: extremo distal a la línea media.

Por tanto, el conjunto de todos estos datos sugiere que los parálogos de la familia Yap podrían controlar los movimientos de convergencia-extensión en embriones de pez medaka, a través de un programa morfogénico establecido en la línea media y en el que los genes diana de la familia CCN, particularmente *ctgfa*, juegan un papel central en la correcta migración de precursores hacia el eje en formación del embrión.



Discusión



1. Introducción

Los diferentes parálogos de la familia Yap y los factores de transcripción de la familia Tead cumplen su función trabajando como un tándem de proteínas reguladoras, pues, por un lado, las proteínas de la familia Yap no pueden interactuar directamente con el genoma, y por el otro, la función de Tead como factor de transcripción depende de la capacidad del dominio de activación carboxi-terminal de Yap1/Taz para reclutar al complejo de la Pol II (Galli et al., 2015; Stein et al., 2015; Vassilev et al., 2001; Zanconato et al., 2015). La naturaleza bipartita del complejo Yap1/Tead se ilustró bien desde los primeros experimentos pioneros con células linfocíticas, en los que la sobreexpresión de Yap aumentaba hasta 300 veces la actividad de Tead (Vassilev et al., 2001). En este proyecto se ha demostrado la existencia de un nuevo parálogo de *yap1*, al que se ha bautizado como *yap1b*, que presenta un dominio de activación transcripcional divergente de los descritos previamente para Yap1 (Sudol, 1994) y para Taz (Kanai et al., 2000).

2. Evolución de parálogos de la familia Yap en teleósteos

Los análisis filogenéticos llevados a cabo en este estudio esclarecen el origen evolutivo de los diferentes parálogos de la familia Yap, así como su distribución entre los diferentes grupos de teleósteos existentes en la naturaleza. Estudios previos han demostrado que Yap1 surgió en metazoos como un acompañante de la función del factor de transcripción Tead, y que su duplicación dio lugar a la aparición de Yap1 y de Taz en el linaje de los vertebrados (Hilman y Gat, 2011). Ambos parálogos comparten una arquitectura muy similar y un rasgo particular, que parece ser específico de vertebrados: la capacidad de ser activados por kinasas de la familia Src mediante fosforilación de su dominio de activación transcripcional (Elbediwy y Thompson 2018; Sudol, 1994). A lo largo de esta tesis se muestra que, durante la duplicación genómica específica de los teleósteos, los dos parálogos ancestrales de vertebrados (Yap y Taz) volvieron a duplicarse para dar lugar a cuatro copias funcionales: *yap1*, *yap1b*, *taza* y *tazb*. De hecho, en el presente estudio mostramos que estos cuatro parálogos están presentes en varias especies ancestrales de teleósteos, como *Elopomorpha* (anguila japonesa) y *Osteoglossomorpha* (pez elefante). Respecto a los parálogos de *taz*, la evolución de los distintos grupos de teleósteos trajo como resultado la eliminación de una primera copia antes de la aparición de las especies del grupo *Othomorpha*, y la eliminación de la segunda exclusivamente en especies del grupo *Percomorpha*. Este hallazgo implica que uno de cada

cuatro vertebrados (correspondientes a las especies del clado de los *Percomorpha*) carece de parálogos de *taz* en su genoma. La historia evolutiva de *yap1* contrasta bastante con la de *taz*. Para empezar, la secuencia de *yap1* apenas ha cambiado desde su aparición y se ha convertido en el parólogo funcional dominante en diferentes especies de teleósteos (Kimelman et al., 2017; Miesfeld et al., 2015; Porazinski et al., 2015). Por el contrario, la copia de *yap1b* se ha perdido en especies del clado *Othomorpha* (al que pertenece el pez cebra) y su dominio de activación transcripcional ha experimentado sustanciales cambios, incluyendo la pérdida de residuos de tirosina, fosforilables por kinasas Src, en euteleósteos (cuyo ejemplo más estudiado es el del pez medaka). La eliminación de estos residuos en *yap1b* supondría, en principio, una mayor dependencia de mecanismos de regulación alternativos, ya sea la propia ruta Hippo o la mecanotransducción de señales, puesto que la fosforilación controlada por proteínas SFK resulta esencial para la activación transcripcional de complejos Yap1/Tea (Calvo et al., 2013; Li et al., 2016). De hecho, cabría preguntarse si la modificación del dominio carboxi-terminal de *yap1b* puede alterar las propiedades de transactivación de esta proteína. Teniendo en cuenta la alta conservación del dominio de unión a Tea de esta proteína y que cumple una función similar a *yap1* según nuestros datos de iDamID-seq, un dominio carboxi-terminal inactivo podría ser compatible con una función dominante negativa. Los datos recabados a lo largo de este proyecto permiten descartar esta posibilidad, puesto que la pérdida de *yap1b* en embriones mutantes de *yap1* en medaka no sólo no genera embriones con menos alteraciones morfológicas sino que intensifica su fenotipo. Por lo tanto, la hipótesis más razonable es que el dominio de activación transcripcional de *yap1b*, al contrario que los parálogos presentes en otros vertebrados, es activo independientemente del control ejercido por kinasas de la familia SFK. Así pues, las propiedades reguladoras de Yap1b se asemejarían más a las de las proteínas de la familia Yap de invertebrados, cuya función tampoco está controlada por parte de kinasas de la familia Src, como es el caso de la proteína Yorkie en *Drosophila* (Elbediwy y Thompson 2018).

3. Análisis de regiones de unión a DNA de proteínas de la familia Yap

Numerosos estudios previos basados en experimentos de ChIP-seq han tratado de identificar los sitios de unión de Yap y Taz o de sus cofactores principales, especialmente Tea, a la cromatina. Estos análisis se han llevado a cabo en células troncales embrionarias de mamífero (Estarás et al., 2017; Lian et al., 2010), en células cancerosas (Galli et al., 2015; Nardone et al., 2017; Stein et al., 2015; Zanonato et al., 2015) o en diferentes tejidos embrionarios (Cebola et

al., 2015; Lin et al., 2017). A pesar de que existen ciertas discrepancias en cuanto al número de regiones identificadas (que van de los 1000 a los 7000 picos, aproximadamente), la mayoría de estos estudios, si no todos, determinan la existencia de un importante solapamiento entre picos asociados a Tead y picos asociados a Yap1/Taz, y demuestran que Tead es el motivo de unión más representativo entre los conjuntos de picos identificados para Yap1 y Taz. Además, estos estudios identifican un importante número de genes, como *amotl2*, *ctgf*, *cyr61* y *ajuba*, como dianas transcripcionales para ambas proteínas.

En este proyecto de tesis se ha utilizado la tecnología iDamID-seq para investigar la identidad de los sitios de unión al genoma más importantes de los diferentes parálogos de Yap, tanto en embriones de pez cebra como de medaka, durante la etapa de gastrulación. La utilización de este método experimental supone dos ventajas añadidas con respecto al uso de técnicas de ChIP-seq. La primera es que no depende del uso de anticuerpos adecuados para los experimentos de inmuno-precipitación, normalmente desarrollados para mamíferos, y que podrían no reconocer los epítomos de las proteínas de teleósteos. La segunda ventaja es que se elimina la necesidad de partir de un gran número de células. Esta última limitación, particularmente, podría ser una de las razones que han impedido el desarrollo de estudios de ocupación de cromatina en las etapas embrionarias más tempranas. Además, la técnica iDamID-seq es especialmente apta para la identificación de las dianas de unión al genoma de proteínas que no se unen directamente al DNA y que requieren de la presencia de un cofactor para llevar a cabo su actividad transcripcional, como es el caso que nos ocupa. El uso de la técnica iDamID-seq se ha extendido para la determinación de sitios de unión al DNA de factores de diversa clase en *Drosophila* (Marshall y Brand 2017; Marshall et al. 2016; Southall et al. 2013) y, más recientemente y una vez resueltos ciertos problemas de toxicidad, en medaka (Gutierrez-Triana et al., 2016) y en mamíferos (Cheetham et al., 2018; Tosti et al., 2018). Numerosas pruebas demuestran que nuestro iDamID-seq ha identificado correctamente los sitios de unión a DNA de los parálogos de la familia Yap durante la embriogénesis temprana en pez cebra y medaka. Por un lado, nuestros análisis por triplicado muestran una alta reproducibilidad en sus resultados, y en todos los casos se identificó la presencia del motivo Tead como uno de los más significativos entre el conjunto de picos obtenido, ya fuese de manera directa empleando la herramienta DREME (como es el caso de los picos obtenidos para las proteínas Yap1 y Yap1b de medaka) o de manera indirecta, asignando el motivo Tead consenso a la herramienta AME (en el caso de los picos obtenidos para Yap1 y Taz en pez cebra). Por otra parte, además de Tead, fue posible identificar un enriquecimiento significativo en numerosos motivos cuya interacción con proteínas de la familia Yap ha quedado demostrada, como es el caso de los factores de transcripción p63, p73

y *tbx5* (Piccolo et al., 2014). Asimismo, muchas de las marcas identificadas, con un solapamiento parcial con regiones abiertas de cromatina, se hallan en las proximidades o justo en la secuencia promotora de genes típicamente asociados a Yap, como *ctgf*, *cyr61*, *amotl2*, *lats2*, *ajuba* o *ccne1* (Lin et al., 2017; Stein et al., 2015; Zanconato et al., 2015).

Estos análisis de iDamID-seq nos han permitido determinar las propiedades de unión a DNA de *yap1b*, el nuevo parálogo identificado de *yap1*. Como ya se había estimado anteriormente según los resultados de conservación de su dominio de unión a Tead, *yap1b* se une a cierto número de regiones de la cromatina, que en gran medida también resultan ser dianas de Yap1. Además, el análisis comparativo llevado a cabo en esta tesis arroja unos resultados que deben servir de punto de partida para investigar el programa transcripcional controlado por los diferentes cofactores que se unen a los distintos parálogos de la familia Yap durante la embriogénesis temprana de los vertebrados. La comparación entre los datos recabados a lo largo de este proyecto y experimentos previos de ChIP-seq sugieren que un amplio porcentaje del programa transcripcional controlado por proteínas de la familia Yap está conservado en vertebrados, independientemente del tipo celular o de la etapa de desarrollo considerada. No obstante, será necesario llevar a cabo análisis más exhaustivos para determinar con exactitud las coincidencias y diferencias presentes en la regulación genética ejercida por estos parálogos en diferentes contextos tisulares y embrionarios.

4. Función cooperativa entre parálogos de la familia Yap durante el desarrollo temprano de teleósteos

Las discrepancias fenotípicas observadas entre mutantes del gen *yap1* en medaka y pez cebra constituyen una de las motivaciones principales de este trabajo. La mutación de dicho gen en peces medaka provoca importantes malformaciones incluso desde las etapas inmediatamente posteriores a la gastrulación, que afectan al desarrollo de numerosos órganos, como el corazón, los ojos, el cerebro y los somitas (Porazinski et al., 2015). Por el contrario, la mutación de este mismo gen en peces cebra apenas provoca algunas alteraciones muy leves, y solo la mutación conjunta de los dos parálogos de la familia Yap, *yap1* y *taz*, genera anomalías anatómicas de importancia similares a las observadas en pez medaka (Kimelman et al., 2017; Miesfeld et al., 2015). Por otra parte, aunque los principales rasgos fenotípicos de los embriones *yap1*^{-/-} de medaka y de los embriones mutantes dobles *yap*^{-/-} *taz*^{-/-} de pez cebra ya han sido descritos previamente, en ninguna de estas publicaciones se había puesto de manifiesto el hecho de que estos mutantes presentan un eje anteroposterior

significativamente más ancho y más corto que el de embriones WT. Este rasgo podría ser indicativo de defectos durante la convergencia-extensión de precursores mesendodérmicos, pues es muy similar al fenotipo asociado a mutaciones en genes descritos previamente implicados en este proceso, como *jak1*, *wnt5* y *wnt11* (Conway et al., 1997; Heisenberg et al., 2000; Kilian et al., 2003). El análisis de estas discrepancias fenotípicas a lo largo de la presente tesis doctoral nos condujo al hallazgo de un tercer parálogo de esta familia de proteínas, *yap1b*, en el genoma de medaka y en el de otros teleósteos menos evolucionados. Es este parálogo y no *taz* el que está presente en el genoma de medaka, al contrario de lo que se venía asumiendo hasta ahora (Porazinski et al., 2015). El análisis comparativo de los perfiles de expresión de cada uno de los parálogos de Yap entre embriones de pez cebra y medaka aporta una explicación a las diferencias fenotípicas observadas entre embriones mutantes de *yap1* en ambas especies. Mientras que los genes *yap1* y *taz* se expresan a unos niveles similares durante las fases de gastrulación y de segmentación en embriones de pez cebra, los niveles de *yap1b* son mucho más bajos que los de *yap1* durante las mismas etapas en medaka. Esta observación sugiere que *taz* podría compensar genéticamente la pérdida de *yap1* en pez cebra, mientras que esta compensación es mucho menor en medaka, al jugar *yap1* un papel más dominante a partir del inicio de la gastrulación. Resulta interesante el hecho de que la contribución materna de *yap1b* sea tres veces más alta que la de *yap1*. Como Yap juega un papel destacado en la activación del genoma cigótico de ratones (Yu et al., 2016), la contribución materna de *yap1b* sugiere posibles funciones en la transición materno-cigótica en medaka. Sin embargo, la pérdida de *yap1b* materno no provoca consecuencia fenotípica alguna, ya que los ejemplares *yap1b*^{-/-} son fértiles y se reproducen normalmente. Por lo tanto, sería necesario llevar a cabo un análisis más exhaustivo para determinar si el bloqueo simultáneo de la contribución materna de *yap1* y de *yap1b* tiene también alguna incidencia sobre la activación cigótica en embriones de medaka, tal y como sucede en ratón.

A pesar de la ausencia de fenotipos destacables en embriones mutantes *yap1b*, los análisis de embriones *yap1*^{-/-} *yap1b*^{-/-} en medaka revelan la implicación de *yap1b* en el control de la morfogénesis de tejidos y de la supervivencia celular durante las etapas tempranas de la embriogénesis. Estas funciones correlacionan con algunos de los términos de GO obtenidos en el análisis de genes asociados a marcas de DamID para ambos genes, como “represión de la apoptosis” o “desarrollo embrionario”. Asimismo, las funciones llevadas a cabo por parálogos de la familia Yap durante la embriogénesis de teleósteos indicadas previamente (Kimelman et al., 2017; Porazinski et al., 2015) también parecen coincidir con estas observaciones. Por otra parte, el rol cooperativo de *yap1b* no depende de la activación controlada por SFKs. Se desconoce el papel exacto que desempeña la fosforilación del dominio de activación

transcripcional de las proteínas Yap a nivel fisiológico, tanto en la embriogénesis como en muchos tejidos adultos. Sin embargo, la activación de Yap vía SFKs parece resultar importante en la regeneración del tejido hepático y de la mucosa intestinal (Taniguchi et al., 2015) y la pérdida de kinasas Fyn-Yes está asociada a defectos en la convergencia-extensión de embriones de pez cebra (Jopling y den Hertog 2005). Tal vez este hecho constituya una alternativa a las diferencias observadas en el nivel de expresión de *yap1* y de *yap1b* a lo largo del desarrollo como posible causa que explicase las anomalías en el proceso de convergencia-extensión observadas en embriones *yap1*^{-/-} de medaka, pues al carecer de un parálogo alternativo a Yap que pueda ser regulado por kinasas de la familia Src/Yes, los defectos producidos por la pérdida de ésta se verían potenciados. Un experimento de sobreexpresión de Yap1b en embriones mutantes *yap1*^{-/-} de medaka permitiría determinar a priori si la ausencia de residuos fosforilables por SFKs es determinante en la función de Yap como un factor de transcripción imprescindible para la correcta gastrulación de embriones de medaka y así determinar cuál de las dos razones expuestas explicaría la diferencia de fenotipos entre mutantes de pez cebra y de pez medaka. En cualquier caso, la particular composición estructural de los parálogos de Yap en euteleósteos, con una única proteína fosforilable por SFKs, abre una oportunidad única en vertebrados para analizar en profundidad la función de este mecanismo de regulación tanto en el desarrollo embrionario como en tejidos adultos.

5. Papel de Yap como agente inductor de migración celular durante la gastrulación de medaka

El análisis de las principales dianas de unión a DNA de los diferentes parálogos de la familia Yap mediante DamID-seq arrojó datos interesantes, que coinciden, al menos parcialmente, con lo expuesto en otros experimentos alternativos llevados a cabo en otros contextos celulares y tisulares. Sin embargo, los resultados obtenidos mediante el empleo de esta técnica no bastan para determinar con exactitud el conjunto de dianas transcripcionales de estos parálogos, pues siempre cabe la posibilidad de que se produzcan falsos positivos o falsos negativos. De hecho, uno de los puntos débiles de esta técnica es que marca el conjunto de regiones metiladas por la proteína de fusión DAM-TF a lo largo de todo el desarrollo hasta la etapa cuyo perfil se desea examinar, por lo que no es de extrañar que algunas marcas identificadas correspondan realmente a regiones cis-reguladoras activas en etapas previas a la gastrulación que dejan de estarlo durante sus últimas fases. Por otra parte, la unión de un factor de transcripción al DNA no implica necesariamente la regulación directa de los genes vecinos, bien porque sea

necesaria la incorporación de factores adicionales o porque el peso regulador del elemento sobre la transcripción del gen sea modesto. Así pues, es necesario recurrir a otros métodos para poder validar estos datos. Por esta razón, se decidió llevar a cabo un análisis comparativo del transcriptoma de embriones de medaka *yap1^{+/+}* y *yap1^{-/-}* de estadio 16 (el mismo que el de los embriones empleados para el iDamID-seq) mediante RNA-seq. Llama la atención el bajo número de genes cuya expresión se reprime o activa en los embriones mutantes. Esto sugiere que, a pesar de la enorme complejidad que rodea a la regulación de la actividad de Yap, éste controla la expresión de un número relativamente reducido de genes durante las primeras etapas del desarrollo embrionario. A pesar de que el número de genes controlados por Yap en medaka resulta un tanto exiguo, el hecho de que casi el 40% de los genes reprimidos en el mutante también estén asociados a picos cercanos de DamID-seq resulta especialmente significativo, habida cuenta de la complejidad que puede alcanzar el programa transcripcional controlado por estos co-reguladores. Así pues, mientras que algunos genes expresados pueden ser regulados directamente por la unión de dicho factor de transcripción a su promotor o enhancer asociado, la expresión de otros genes puede ser el resultado indirecto de la regulación llevada a cabo a través de la expresión de un tercero que sí esté controlado directamente por el factor de transcripción en estudio.

Al profundizar en la función llevada a cabo por el conjunto de genes reprimidos en los embriones mutantes y, por tanto, cuya expresión depende de la presencia de Yap en el núcleo, encontramos básicamente cinco tipos de genes: genes que codifican para componentes de la matriz extracelular o involucrados en procesos de adhesión (*ctgfa*, *cyr61*, *frem1a*, *itgb4*, *itga3a*, *itga3b*, *col9a2*...), genes implicados en la reorganización del citoesqueleto y en la migración celular (*marcksl1a*, *marcksl1b*, *acta1*, *myo9ab*, *pls3*, *rnd1a*, *vcl*...), genes pertenecientes a la ruta Hippo o a otros inputs reguladores de la actividad de Yap/Taz (*yap1*, *src*, *tead4*, *amotl2a*, *amotl2b*, *lats2*...), genes implicados en apoptosis (*boka*, *pea15*, *pim3*...), y genes que codifican para proteínas pertenecientes a otras rutas de señalización o a otros factores de transcripción (*cxcl14*, *gata2*, *tp63*, *shisa2*, *tgfb3*, *foxo1*...). Esta clasificación parece refrendada en la naturaleza de buena parte de los términos de GO obtenidos tras el correspondiente análisis (*adhesión celular*, *transducción de señales*, *comunicación celular*, *organización del citoesqueleto*...). La expresión, controlada por Yap, de genes codificantes de proteínas implicadas en la formación de adhesiones focales como vinculinas (*vcl*), protocadherinas (*pcdh12*, *pcdh19* y *pcdh20*) y diversas integrinas, y de genes implicados en aumentar la estabilidad del citoesqueleto ya había sido notificada previamente en células humanas cultivadas (Nardone et al., 2017), si bien no se han hallado indicios que demuestren una mayor expresión de proteínas involucradas activamente en una reorganización de la

matriz extracelular, tales como diversas metaloproteinasas. Por otra parte, las proteínas MARCKS, cuyo nivel de expresión se reduce muy significativamente en ausencia de Yap, juegan un papel muy importante durante la gastrulación, puesto que controlan procesos de adhesión celular y de reorganización del citoesqueleto de actina cortical, que resultan ser claves para una correcta convergencia-extensión (Ilioka et al., 2004).

Curiosamente, en la lista de genes regulados por Yap en el experimento de RNA-seq hay pocas referencias a genes codificantes de proteínas que participan en algunas de las rutas de señalización más importantes de la gastrulación. Entre los ejemplos más destacados se hallan los genes *fzd8*, que codifica para uno de los receptores Frizzled que captan a los ligandos Wnt y cuyo nivel de expresión se reduce a la mitad en los mutantes de *yap1*, y *shisa2*, que participa en la especificación de los precursores del mesodermo presomítico en ratones inhibiendo parcialmente las señales FGF y Wnt (Furushima et al., 2007) y que se expresa a una tasa tres veces inferior en ausencia de Yap1. También parece existir relación entre la actividad de Yap y la expresión de los genes *tgfb3*, *bmp4* y *bmp7a*. No obstante, si Yap ejerciera una influencia verdaderamente destacable sobre elementos de otras rutas de señalización morfogénicas, sería esperable un aumento significativo en el número de genes reprimidos en los embriones mutantes. Por otra parte, Yap regula su propia actividad transcripcional a través de la expresión de genes que codifican para proteínas que forman parte de mecanismos que controlan su actividad. Este es el caso de *lats2*, *amotl2a* y *amotl2b*, que ejercen una influencia inhibitoria sobre Yap, y de otros genes como *tead4* o *src*, cuya expresión favorecería la actividad transcripcional de esta proteína. También destaca el caso del gen *vgl13*, que se encuentra sobreexpresado en embriones mutantes *yap1*^{-/-}. Esta proteína forma parte de la familia VGLL, que se cree que compite con Yap por su unión a Tead, por lo que dicha proteína ostentaría un papel inhibidor de la función de Yap, si bien este extremo sólo ha podido ser vinculado específicamente a la proteína VGLL4 (Deng y Fang, 2018). Por último, Yap1 induce su propia expresión en embriones WT.

El conjunto de observaciones anotadas tras este análisis invitan a pensar en Yap1 como una proteína que controla procesos de migración celular durante la gastrulación al promover la expresión de genes involucrados en la formación de complejos de adhesión célula-ECM y en la reorganización del citoesqueleto, fundamentalmente, más que a controlar rutas de señalización clásicas implicadas en esta etapa del desarrollo.

6. Yap1 dirige la migración de células hacia la línea media a través del control de la expresión de *ctgfa*

Al analizar el patrón de expresión de los principales genes controlados por Yap según lo deducido de los resultados del RNA-seq, descubrimos que muchos de ellos, incluido *yap1*, se expresan con especial intensidad a lo largo de la línea media en embriones de pez cebra durante las etapas inmediatamente posteriores a la gastrulación.

Uno de estos genes es *ctgfa*, una diana conocida de Yap (Zhao et al., 2008) cuya expresión es más de siete veces menor en ausencia de Yap1 durante la gastrulación. Junto con su parólogo *cyr61*, cuya expresión también se reduce en ausencia de Yap, *ctgfa* forma parte de la familia de proteínas CCN, un conjunto de proteínas que participan en el ensamblaje de la matriz extracelular y que son capaces de controlar procesos de adhesión celular, migración y quimiotaxis. Estas propiedades sugieren que ambas proteínas podrían desempeñar un rol importante durante la gastrulación de vertebrados. De hecho, se ha demostrado que, además de controlar la integridad de la matriz extracelular y la formación de adhesiones célula-célula y célula-ECM, Cyr61 tiene un papel activo en gastrulación. Así, se ha visto que, en embriones de *Xenopus*, Cyr61 regula las rutas de señalización Wnt y BMP, y que en ausencia de esta proteína los embriones presentan un eje antero-posterior más corto e importantes malformaciones en la región posterior (Latinkic et al., 2003). Por otra parte, la expresión de *ctgfa* desempeña una función importante en la correcta migración de células precursoras cardíacas hacia la línea media ya que mantiene la integridad estructural del endodermo, al promover el acoplamiento de sus células a la matriz extracelular (Fukui et al., 2014). También favorece la regeneración de la médula espinal en peces cebra (Mokalled et al., 2016).

Para determinar la posible función de *ctgfa* y de *cyr61* en la gastrulación de embriones de medaka, se llevaron a cabo ensayos funcionales inyectando morfolinós específicos contra ambas proteínas. Tanto los embriones morfantes de *ctgfa* como los morfantes dobles de *ctgfa* y *cyr61* presentan un eje antero-posterior más corto que se va haciendo especialmente patente a medida que el desarrollo progresa, lo que evidencia problemas de gastrulación similares a los observados en embriones *yap1*^{-/-}. Sin embargo, los morfantes de *cyr61* no muestran un acortamiento significativo, lo que sugiere que *ctgfa* desempeña una función dominante durante la gastrulación de medaka. Asimismo, la sobreexpresión asimétrica de la proteína Ctgfa de pez cebra en embriones de medaka provoca importantes defectos de gastrulación, tanto de epibolia como de convergencia-extensión. Tanto las funciones asociadas a las proteínas de la familia CCN como su localización en la línea media sugieren una función quimiotáctica en el contexto de la gastrulación. El uso de esferas de agarosa que establezcan

gradientes de esta proteína o el desarrollo de experimentos de transferencia celular de embriones inyectados con esta proteína a embriones WT nos permitiría en el futuro obtener unos resultados concluyentes sobre las propiedades quimioattractoras de estas proteínas.

Para complementar los resultados de estos análisis funcionales, se está trabajando en el desarrollo de líneas mutantes estables de *ctgfa* y *cyr61* en medaka. Hasta la fecha, ya ha sido posible establecer una línea mutante estable para *ctgfa*, pero aún no se ha podido determinar con exactitud su fenotipo. Por otra parte, se está intentando establecer una línea mutante estable del gen *cyr61*. La inyección ventral de Cyr61 en *Xenopus* induce la formación de dobles ejes y un cierre tardío del blastoporo (Latinkic et al., 2003). Teniendo en cuenta este dato y los resultados recabados a lo largo de esta tesis, no es descartable que la mutación de alguno de estos dos genes, o bien la mutación simultánea de ambos, termine generando fenotipos similares. Por otra parte, el uso de la tecnología CRISPR contra el gen *ctgfa* en medaka genera embriones de generación F0 con evidentes problemas de desarrollo, con un marcado acortamiento de su eje antero-posterior, lo que constituye una prueba más de la función de este gen en la gastrulación de estos embriones. Así pues, el establecimiento de líneas mutantes *cyr61*^{-/-} y *ctgfa*^{-/-} e incluso el de una línea mutante doble *ctgfa*^{-/-} *cyr61*^{-/-}, en combinación con los experimentos de sobreexpresión de ambas proteínas anteriormente planteados, se antoja imprescindible para comprender con mayor exactitud la función de ambos genes, cuya expresión depende de Yap1, durante la gastrulación.

Por último, quisimos profundizar aún más en la labor de las proteínas CCN como moléculas implicadas en la gastrulación de peces medaka, así que se llevaron a cabo experimentos de *imaging* para determinar la posible existencia de anomalías durante la formación del eje embrionario. Para ello, se analizaron embriones control, inyectados con CRISPR contra *ctgfa* y tratados con morfolinós contra *ctgfa* y *cyr61*. Los resultados obtenidos sugieren la existencia de anomalías en los procesos de epibolia y convergencia-extensión.

El análisis detallado de trayectorias celulares de embriones WT, CRISPantes y morfantes en regiones posteriores y anteriores del eje embrionario en formación también arroja pistas que determinan la existencia de anomalías en la gastrulación en estas dos últimas condiciones. En la parte posterior del eje, las células de los embriones WT, además de seguir una trayectoria constante hacia el polo vegetal al tiempo que migran de manera regular hacia la línea media, presentan un mayor desplazamiento cuadrático medio que las células CRISPantes y morfantes, así como un mayor cociente de direccionalidad. Teniendo en cuenta que las tres poblaciones celulares se desplazan prácticamente a la misma velocidad, estos datos sugieren que la trayectoria de las células de los embriones WT refleja una migración dirigida hacia la parte posterior de la línea media. Por el contrario, la trayectoria de las células de los embriones

CRISPantes y morfantes examinados es mucho más errática, y su migración direccionada hacia la región posterior de la línea media se ve interferida.

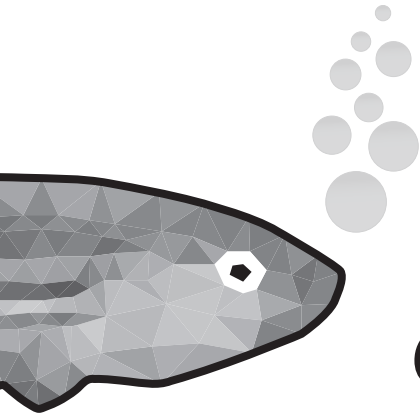
Así pues, el conjunto de los datos recabados de experimentos funcionales, hibridaciones in situ y experimentos de imaging sugiere que la función de *cyr61* y especialmente de *ctgfa* es necesaria para la correcta especificación del eje del embrión en el pez medaka. Los datos previos de la literatura indican que esta función puede extenderse a otros modelos vertebrados.

7. Observaciones finales

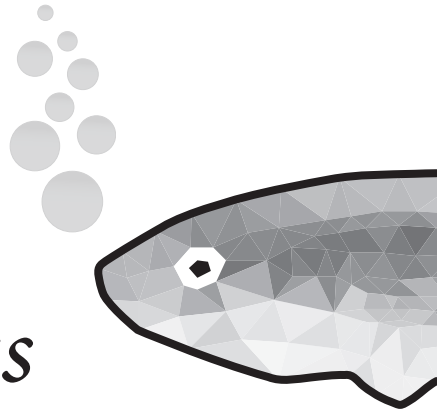
En resumen, en esta tesis doctoral se ha identificado un nuevo parólogo de la familia Yap, que hemos bautizado como Yap1b, el cual es estructuralmente diferente de los parálogos conocidos Yap1 y Taz, al carecer de residuos de tirosina y fenilalanina en su dominio de activación transcripcional, lo que la convierte en una proteína que escapa al control regulador ejercido por kinasas SFK. Este nuevo parólogo surgió tras la duplicación genómica de teleosteos y, a pesar de haber desaparecido en diversos grupos de este clado, como *Otomorpha*, sí está presente en especies del género *Percomorpha*, reemplazando a Taz como parólogo de Yap en una cuarta parte de los vertebrados aún existentes en la actualidad. A pesar de su dominio de transactivación divergente, Yap1 y Yap1b conservan mayor homología en su dominio de unión a Tead que Yap1 y Taz, por lo que no es de extrañar que cumplan funciones similares durante la gastrulación de medaka. El conjunto de datos recabados de la observación fenotípica de mutantes *yap1*^{-/-} de medaka y de experimentos de DamID, RNA-seq, hibridación in situ, ensayos funcionales y experimentos de imaging sugieren que Yap1 controla procesos de migración celular durante la gastrulación, a través de la expresión de genes implicados en el establecimiento de adhesiones celulares y en la reorganización del citoesqueleto.

Sin embargo, aún persisten ciertas cuestiones que por el momento no se han podido resolver: ¿la diferencia fenotípica entre mutantes de *yap1* en pez cebra y medaka se debe a la ausencia de compensación de *yap1b* en medaka durante la gastrulación o a que este parólogo carece de un dominio fosforilable por kinasas SFK, cuya relación con la gastrulación en pez cebra ya ha sido probada? ¿Verdaderamente se producen defectos de convergencia-extensión en embriones mutantes de *yap1* y de *ctgfa* en medaka? ¿La activación de *yap1* durante la gastrulación se debe a la influencia de la tensión morfogenética que se produce en el embrión

durante esta etapa del desarrollo? Futuros experimentos deberán de arrojar algo de luz para tratar de dar una respuesta razonable a estas preguntas.



*Conclusiones /
Concluding remarks*

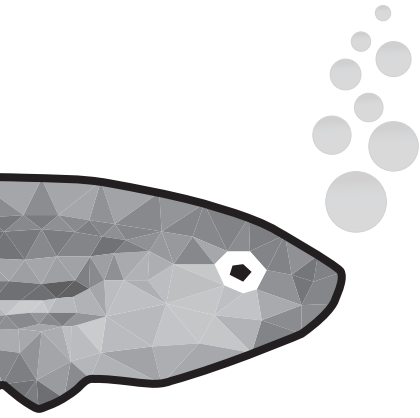


1. Se ha descubierto un tercer parólogo de la familia Yap, no identificado hasta la fecha. Es este tercer parólogo, bautizado como *yap1b*, y no *taz*, el que acompaña a *yap1* en el clado *Percomorpha*, que constituye la cuarta parte de todas las especies actuales de vertebrados.
2. El dominio de activación transcripcional de Yap1b carece de aminoácidos de tirosina y fenilalanina por lo que, a diferencia de Yap y de Taz, no puede ser fosforilado por kinasas de la familia Src.
3. Yap1 y Yap1b conservan una similitud en su dominio de unión a factores de transcripción Tead mayor que la existente entre Yap1 y Taz.
4. Yap1b surgió como el resultado de una duplicación del gen *yap1* hace aproximadamente 350 millones de años. Es posible que adquiriera su estructura divergente tras un proceso de neo-funcionalización.
5. Los genes *yap1* y *yap1b* se expresan de manera ubicua en pez medaka durante la gastrulación. A diferencia de *taz*, la contribución materna de *yap1b* es muy alta, pero su nivel de expresión disminuye drásticamente tras la activación del genoma cigótico. Ésta podría ser una explicación de por qué el fenotipo mutante de *yap1* en medaka es más severo que en pez cebra.
6. Yap1 y Yap1b tienden a incorporarse a regiones abiertas de cromatina cercanas al sitio de inicio de la transcripción de genes implicados en proliferación, diferenciación y apoptosis a través de su unión a factores de transcripción de la familia TEAD, preferentemente.
7. La mutación del gen *yap1* en medaka y la mutación simultánea de *yap1* y de *taz* en pez cebra genera embriones con un eje antero-posterior más corto y más ancho, lo que sugiere la existencia de problemas de convergencia-extensión durante la etapa de gastrulación.
8. Los genes *yap1* y *yap1b* mantienen una relación epistásica negativa. Una mutación en el gen *yap1b* de medaka no provoca efectos fenotípicos obvios, pero la mutación simultánea de ambos genes agrava los defectos observados en mutantes del gen *yap1*, incrementándose significativamente los niveles de apoptosis celular de sus tejidos.

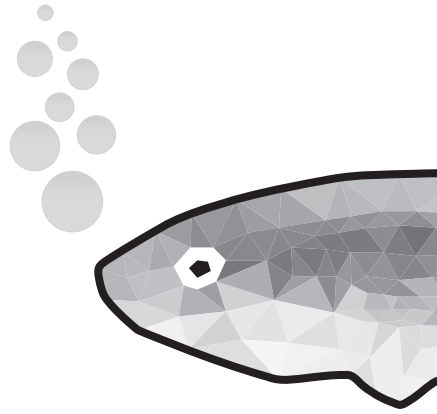
9. Durante la gastrulación de embriones de pez medaka, Yap1 controla la expresión tanto de genes que participan en la regulación de su propia actividad como de genes implicados en la reorganización de las interacciones célula-ECM y del citoesqueleto.
10. Yap1 contribuye decisivamente al establecimiento del eje principal del embrión, regulando los movimientos de convergencia-extensión de las células precursoras a través de la expresión de genes asociados a procesos de adhesión y migración celular, especialmente proteínas de la familia CCN.

1. A new member of the Yap protein family has been identified. This new paralog, called *yap1b*, and not *taz*, escort *yap1* in *Percomorpha*, a group of teleosts representing a quarter of the living vertebrates.
2. Yap1b's transactivation domain lacks tyrosine and phenylalanine residues. Therefore, unlike Yap and Taz, it cannot be phosphorylated by Src/Yes kinases.
3. The Yap1b's Tead binding domain shows higher similarity to the equivalent Yap1's domain than to that of Taz.
4. *yap1b* emerged as a result of the duplication of the *yap1* gene about 350 million years ago. It may have acquired its divergent structure after a neo-functionalization process.
5. *yap1* and *yap1b* are expressed ubiquitously in medaka during gastrulation. Unlike *taz*, there is a high maternal contribution of *yap1b*, but its expression level is drastically reduced after the zygotic genome activation. This could explain why the phenotype of *yap1* mutant medaka embryos is more severe than in zebrafish embryos.
6. Through preferential interaction with Tead transcriptional factors, Yap1 and Yap1b bind open chromatin regions close to the TSS of genes often involved in apoptosis, proliferation and cell differentiation.
7. The mutation of *yap1* gene in medaka and the double mutation of *yap1* and *taz* genes in zebrafish give rise to embryos with a shorter and wider anterior-posterior axis. This evidence suggests an abnormal convergent-extension during gastrulation.
8. There is a negative epistatic relationship between *yap1* and *yap1b*. Mutations in the *yap1b* gene do not show any obvious phenotype. However, the simultaneous mutation of both genes aggravates the anatomical defects observed in *yap1*^{-/-} embryos, showing a significant increase in the apoptotic levels.
9. During gastrulation in medaka, Yap1 regulates the expression of genes involved in the establishment of cell-ECM interactions and in cytoskeleton organization, as well as genes that play an important role controlling its own activity.

10. Yap1 contributes decisively to the body axis formation of medaka embryos as well as to a correct convergent-extension of cell progenitors. These functions are exerted by regulating the expression of genes involved in cell adhesion and migration, especially genes that encode for CCN proteins.



*Material
adicional*



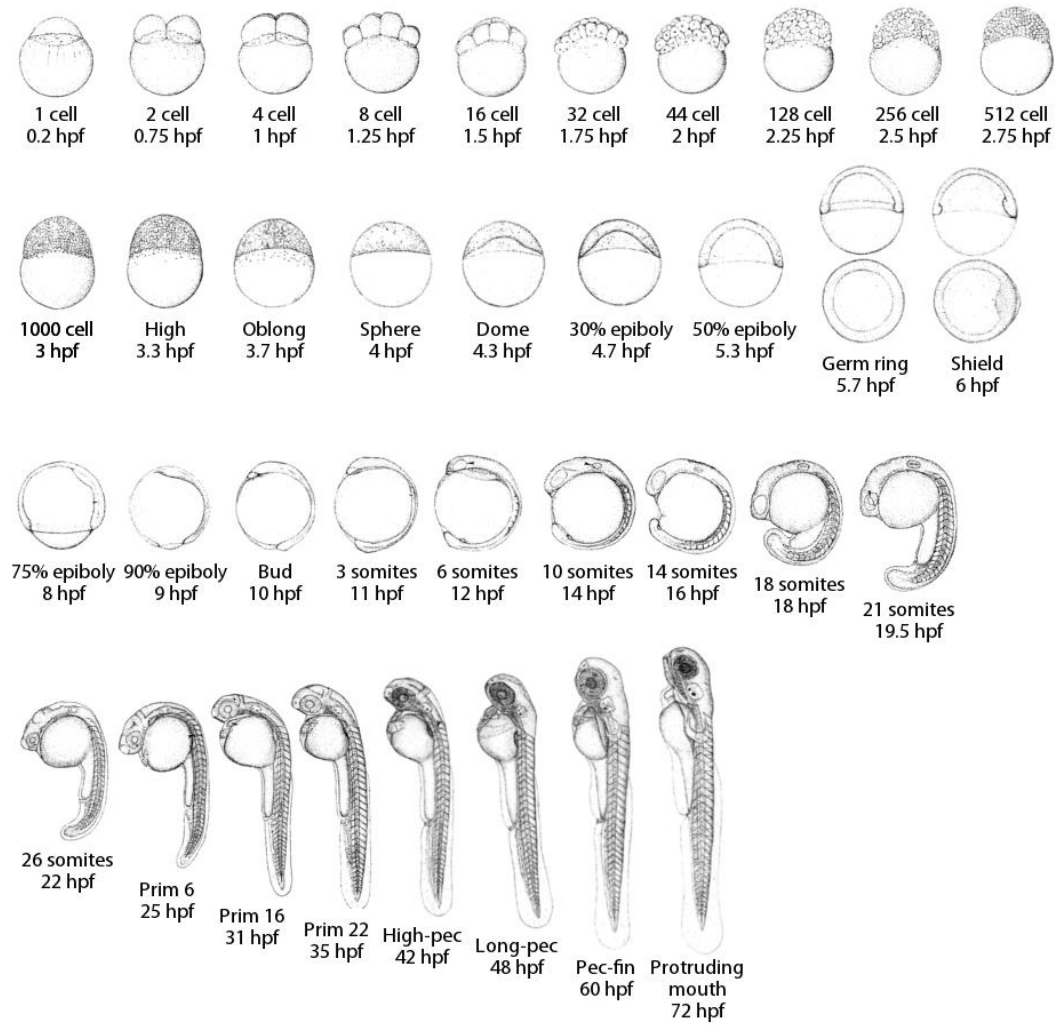


Figura Suplementaria 1 – Etapas de desarrollo embrionario de pez cebra. Representación gráfica de los rasgos morfológicos de cada una de las etapas de desarrollo de embriones de pez cebra (*Danio rerio*). Junto con su correspondiente nombre, aparece representado, en horas post-fecundación (hpf), el periodo de tiempo requerido para alcanzar cada etapa, siempre que los embriones se mantengan a una temperatura constante de 28.5°C, según las pautas establecidas. Figura modificada de Kimmel et al., 1995.

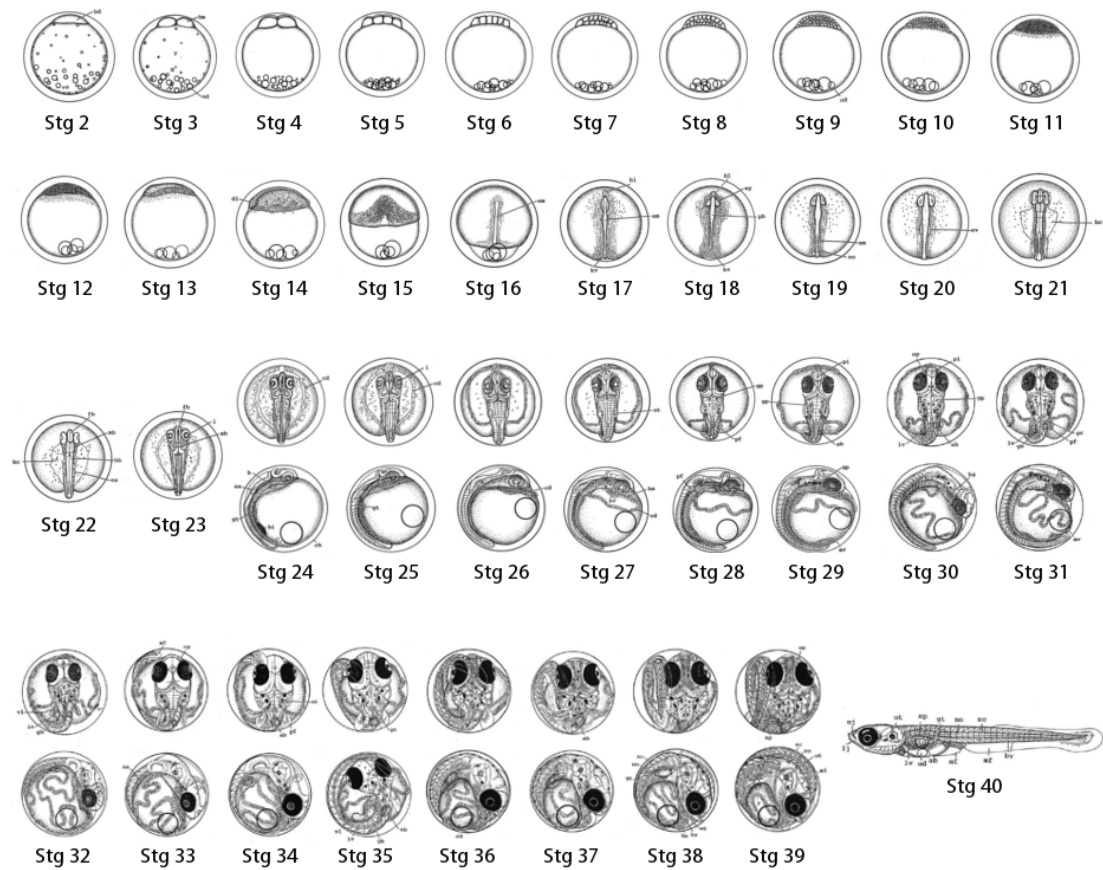


Figura Suplementaria 2 – Etapas de desarrollo embrionario de pez medaka. Representación gráfica de los rasgos morfológicos de cada una de las etapas de desarrollo de embriones de pez medaka (*Oryzias latipes*). Figura modificada de Iwamatsu, 2004.

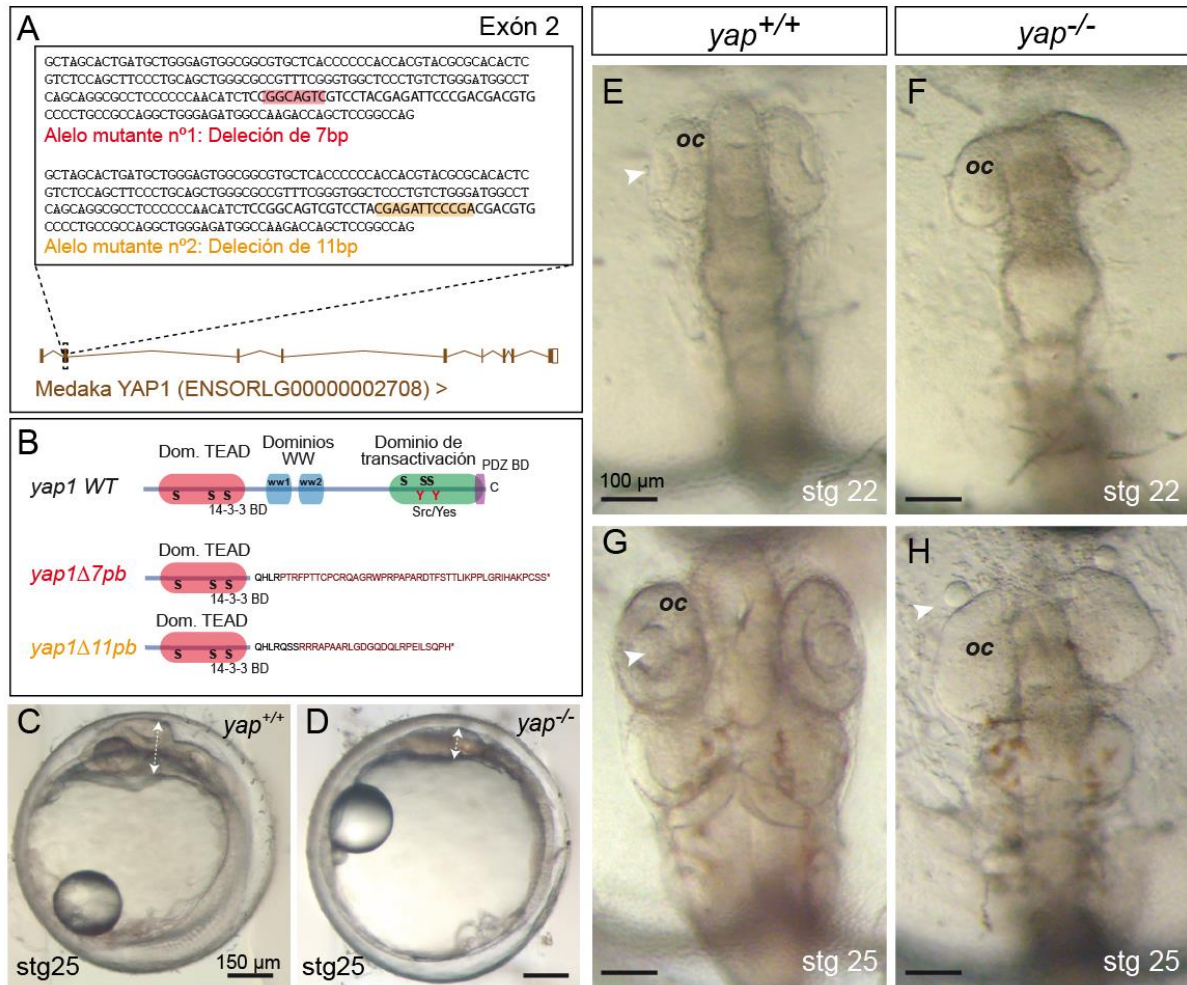


Figura Suplementaria 3 – Generación de alelos mutantes en el gen *yap1* de medaka mediante tecnología CRISPR-Cas9. A) Representación esquemática de las delecciones de 7 y 11 pb generadas sobre el exón 2 del gen *yap1* en medaka. La secuencia en negrita corresponde a la región diana de dos sgRNAs parcialmente solapantes (vea *Materiales y métodos*). Las regiones sombreadas en rojo y naranja corresponden a los nucleótidos delecionados. **B)** Estructura proteica estimada para los alelos WT y mutantes de *yap1*. En ambos casos, las mutaciones generadas producen la pérdida de la pauta abierta de lectura del gen y, por tanto, la aparición de un codón de STOP prematuro y la síntesis de una versión truncada de la proteína. **C-H)** Análisis fenotípico de embriones WT (C, E, G) y mutantes (D, F, H) en estadios de desarrollo 22 (E-F) y 25 (C, D, G, H). Los embriones mutantes muestran malformaciones muy características, como un aplastamiento severo de las copas ópticas y la desalineación de las lentes (flecha de la figura H). Leyenda: oc: copa óptica.

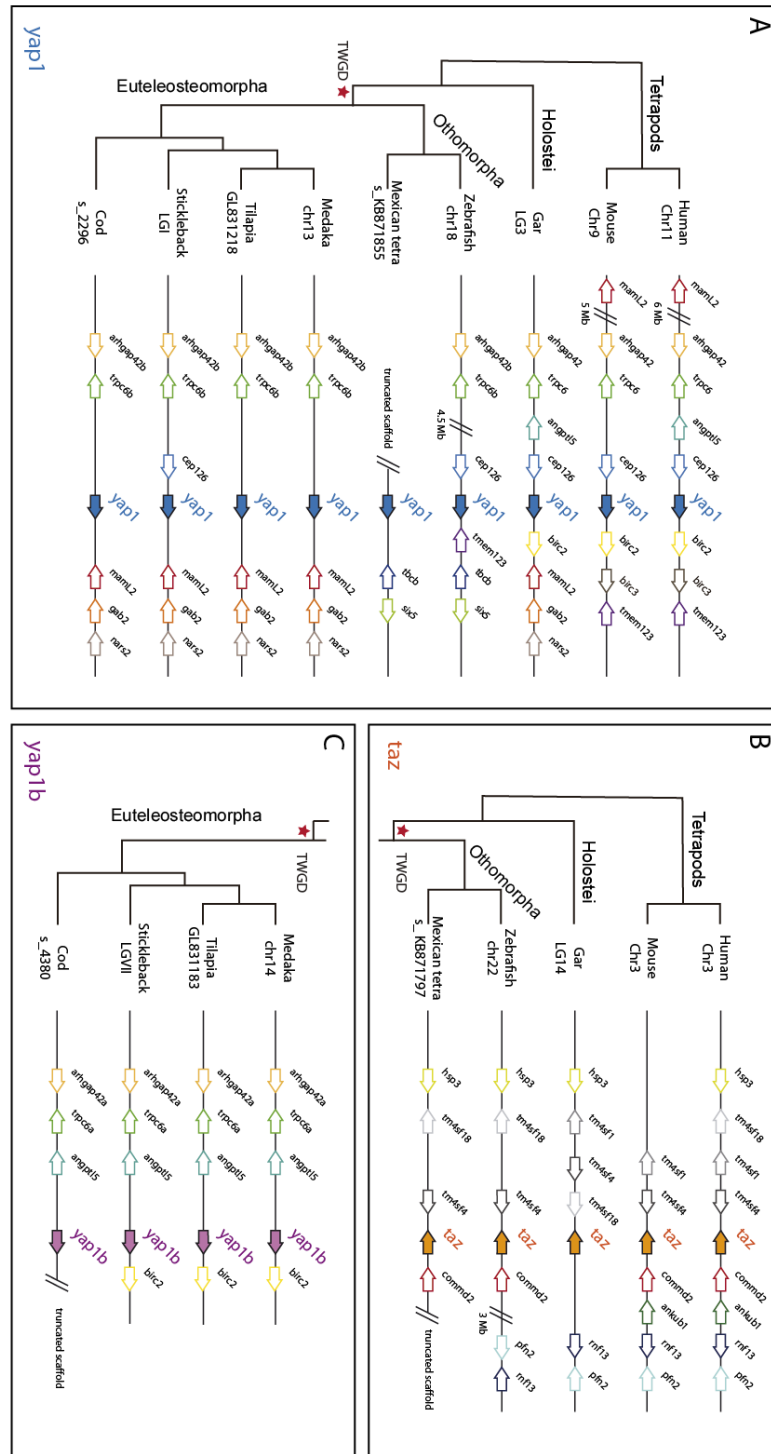


Figura Suplementaria 4 – Análisis sinténico de parálogos de la familia Yap. A) Árbol filogenético que muestra la disposición cromosómica de los genes cercanos al locus del gen *yap1* en las 7 especies de teleosteos y 2 de tetrapodos consideradas como representativas y analizadas en la Figura 9. Obsérvese la conservación del bloque *arhgap42/trcp6* a lo largo de las especies propuestas. B) Árbol filogenético que muestra la disposición cromosómica de los genes cercanos al gen *taz* a lo largo de la evolución de varios grupos de vertebrados. Nótese la alta conservación de los genes *hsp3* y *comm2*. C) La conservación de *arhgap42a* y *trcp6a* como genes cercanos a *yap1b* demuestra que este gen surgió como una duplicación de la proteína *yap1* ancestral. El código de flechas coloreadas indica tanto la orientación como los equivalentes ortólogos de cada gen para cada una de las especies examinadas. También se indica el cromosoma al que pertenece cada bloque sinténico para cada especie. Leyenda: TWGD: Duplicación genómica de teleosteos.

A

Proteína total	1	2	3	4	5	6	7	8	9	10	11	12	13
Yap1-Human_	100	78.0	74.0	70.5	67.4	72.0	36.1	42.1	28.2	38.1	34.0	42.4	37.4
Yap1-Gar_	78.0	100	83.3	79.8	75.2	82.3	40.0	45.0	30.5	28.0	38.5	39.0	41.3
Yap1-Zebrafish_	74.0	83.3	100	81.4	79.0	86.0	39.0	40.2	33.0	36.1	39.5	38.0	38.7
Yap1-MedakaGB_	70.5	79.8	81.4	100	81.0	82.3	37.1	42.3	32.3	31.1	35.3	39.0	44.2
Yap1-Cod_	67.4	75.2	79.0	81.0	100	80.7	37.1	41.1	35.0	34.5	42.4	38.7	
Yap1-Pike_	72.0	82.3	86.0	82.3	80.7	100	39.0	42.5	30.5	37.0	35.1	45.1	38.5
Taz-Human_	38.1	40.0	39.0	37.1	37.1	39.0	100	65.0	69.3	58.1	24.4	27.1	21.1
Taz-Gar_	42.1	45.0	40.2	42.3	41.1	42.5	65.0	100	62.0	72.0	27.0	27.8	30.5
Taz-Zebrafish_	28.2	30.5	33.0	32.3	27.5	30.5	69.3	62.0	100	65.1	22.3	27.8	24.5
Taz-Pike_	38.1	28.0	30.7	31.1	30.0	37.0	58.1	72.0	65.1	100	25.2	28.1	21.3
Yap1b-MedakaGB_	34.0	38.5	39.5	39.3	34.5	35.1	24.4	27.0	22.3	25.2	100	63.4	46.0
Yap1b-Cod_	42.4	39.0	38.0	39.0	42.4	45.1	27.1	27.8	27.8	28.1	63.4	100	53.2
Yap1b-Pike_	37.4	41.3	39.7	44.2	38.7	39.5	21.1	30.5	24.5	21.3	46.0	53.2	100

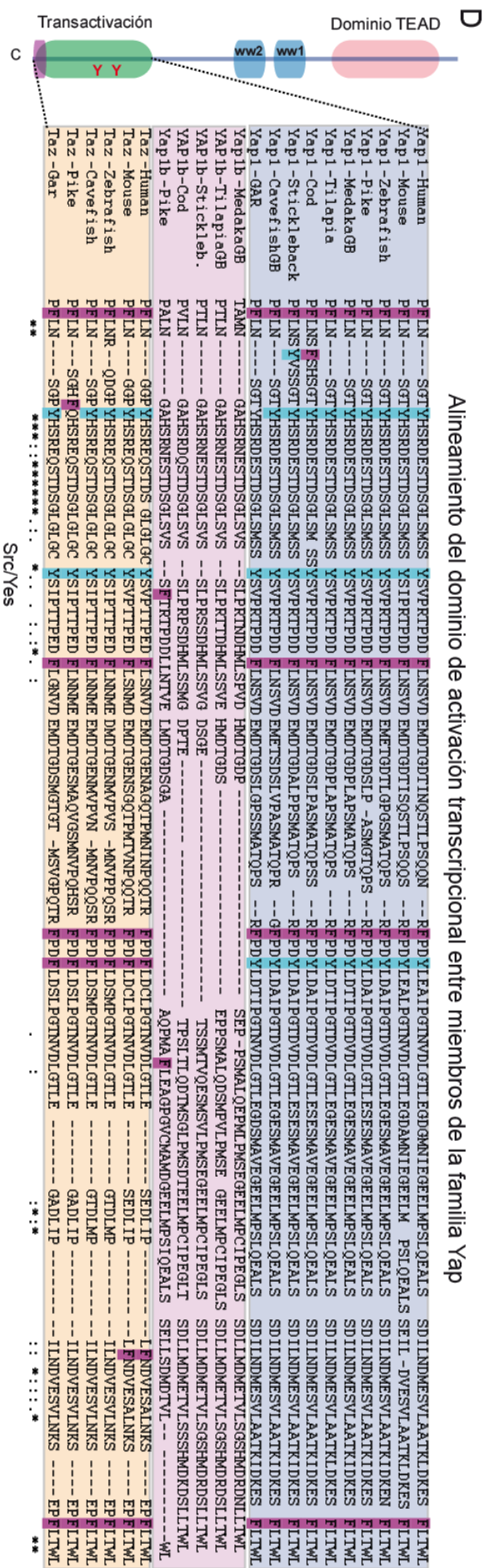
B

Domino Tead	1	2	3	4	5	6	7	8	9	10	11	12	13
Yap1-Human_	100	88.2	89.5	90.0	89.4	89.0	49.4	48.1	46.2	53.0	59.3	67.4	65.1
Yap1-Gar_	88.2	100	95.0	92.5	92.4	94.7	44.7	45.7	41.0	48.2	57.0	67.1	66.4
Yap1-Zebrafish_	89.5	95.0	100	99.2	91.0	95.3	45.9	44.4	40.2	48.4	58.1	68.0	65.1
Yap1-MedakaGB_	89.0	92.4	98.2	100	89.2	89.4	47.1	46.0	47.4	53.0	63.5	70.0	68.2
Yap1-Cod_	88.4	92.4	91.0	89.2	100	94.2	48.2	46.0	44.9	51.8	61.0	70.0	68.8
Yap1-Pike_	89.5	94.7	95.3	99.4	94.2	100	47.1	44.4	47.4	49.4	59.3	68.0	67.4
Taz-Human_	49.4	44.7	45.0	47.1	48.2	47.1	100	76.5	70.5	77.1	94.1	40.0	41.2
Taz-Gar_	48.1	45.7	44.4	46.0	46.0	44.4	76.5	100	73.1	85.2	37.0	40.7	42.0
Taz-Zebrafish_	46.2	41.0	46.2	47.4	44.0	47.4	70.5	73.1	100	75.0	38.5	41.0	41.0
Taz-Pike_	53.0	48.2	49.4	53.0	51.8	49.4	77.1	85.2	75.0	100	38.0	44.0	44.0
Yap1b-MedakaGB_	59.3	57.0	58.1	63.5	61.0	59.3	94.1	37.0	38.5	38.0	100	77.0	67.8
Yap1b-Cod_	67.4	67.1	68.0	70.0	70.0	69.0	40.0	40.7	41.0	44.0	77.0	100	66.2
Yap1b-Pike_	65.1	69.4	65.1	68.2	69.0	67.4	41.2	42.0	41.0	44.0	67.8	66.2	100

C

Dom. Transa.	1	2	3	4	5	6	7	8	9	10	11	12	13
Yap1-Human_	100	89.0	84.4	83.5	82.0	83.3	51.0	50.3	47.0	42.9	44.0	36.8	46.7
Yap1-Gar_	89.0	100	94.3	94.2	87.7	92.9	49.0	50.2	51.4	45.7	48.0	36.0	51.1
Yap1-Zebrafish_	84.4	94.3	100	93.4	87.7	92.9	47.1	57.3	49.5	41.0	48.0	36.8	48.9
Yap1-MedakaGB_	83.5	94.2	93.4	100	90.0	95.0	47.1	54.4	48.0	50.5	46.0	37.0	50.0
Yap1-Cod_	82.0	87.7	87.7	90.0	##	94.2	44.2	53.4	50.5	42.9	46.0	40.0	47.8
Yap1-Pike_	83.5	92.0	92.0	95.0	94.2	100	51.0	47.0	47.0	46.5	45.0	40.0	48.9
Taz-Human_	51.0	49.0	47.1	47.1	44.2	51.0	100	79.0	72.1	74.0	23.5	23.2	27.8
Taz-Gar_	50.3	50.2	57.3	54.4	53.4	47.0	79.0	100	80.0	80.0	23.5	21.1	28.7
Taz-Zebrafish_	47.0	51.4	48.5	48.0	50.5	47.0	72.1	80.0	100	80.0	22.4	21.1	25.0
Taz-Pike_	42.0	45.7	41.0	50.5	42.0	49.5	74.0	80.0	80.0	100	19.4	20.0	28.7
Yap1b-MedakaGB_	44.0	48.0	48.0	46.0	46.0	45.0	23.5	23.5	22.4	19.4	100	68.5	52.2
Yap1b-Cod_	36.8	36.0	36.0	37.0	40.0	40.0	23.2	21.1	21.1	20.0	68.5	100	41.1
Yap1b-Pike_	46.7	51.1	48.0	50.0	47.8	48.0	27.8	26.7	25.0	28.7	52.2	41.1	100

Alineamiento del dominio de activación transcripcional entre miembros de la familia Yap



A

Alineamiento del extremo N-terminal del dominio de transactivación de proteínas Yap

Anja_Yap1	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEM
Medaka_Yap1	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEM
Paking_Yap1	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEM
Gar_Yap	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEM
Tilapia_Yap1	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEM
Pike_Yap1	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEM
Zebrafish_Yap1	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEME
Herring_Yap1	-GTYHSRDESTDSGLSMSSYS-VPRTPD ^{FLNSV} DEME
Paking_Yap1b	-GTYHSRDESTDSGLSVSSYS-VPRTPD ^{LLNSV} EEME
Anja_Yap1b	-GNFHSRDESTDSGLSVSSYS-IPRTPD ^{FLNSV} DEM
Medaka_Yap1b	-NGAHSRNESTDSGLSVS--S-LPRTNDHMLSPVDHMD
Tilapia_Yap1b	-NGAHSRNESTDSGLSVS--S-LPRTNDHMLSSVEHMD
Pike_Yap1b	-NGAHSRNESTDSGLSVS--S-FTRTPD ^{LLNTVEL} MD
Anja_Tazb*	-GPFHSREQSTD ^{SGLGLGCYS} -IPTTPED ^{FLNNV} DEM
Gar_Taz	-GPFHSREQSTD ^{SGLGLGCYS} -IPTTPED ^{FLGNV} DEM
Zebrafish_Taz	-GPFHSREQSTD ^{SGLGLGCYS} -IPTTPED ^{FLNNME} MD
Pike_Taz	GHEQHSREQSTD ^{SGLGLGCYS} -IPTTPED ^{FLNNME} MD
Anja_Taza*	-GPFHSREQSTD ^{SGLGLGCYS} -IPTTPED ^{ILNNVE} MD
Paking_Taz1*	-GPFHSREQSTD ^{SGLGLGCYS} -VPTTPED ^{FLNNLE} MD
Herring_Taz	-GPFHSREQSTD ^{SGLGLNCTNCPTTPED} FLNNMEMD
Paking_Taz2*	-GPHSRDKSTD ^{SGLGLGCCS} -HPTTPE ^{GLS} -MDEM

:**.:. : * : * . : * : *

Src/Yes

B

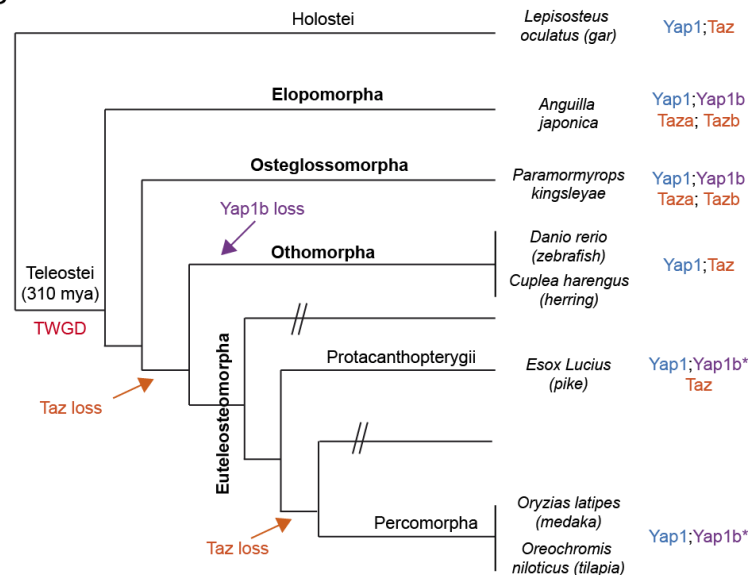


Figura Suplementaria 6 – Historia evolutiva de la familia Yap en teleósteos. A) Alineamiento múltiple de secuencias de la región N-terminal del dominio de transactivación de parálogos de la familia Yap en diferentes especies de teleósteos. Obsérvese que la eliminación o sustitución, en las secuencias de Yap1b, de residuos conservados de tirosina (púrpura) y fenilalanina (azul claro) ocurre en especies del grupo *Euteleostomorpha*, pero no en especies de *Elopomorpha* ni del grupo *Osteoglossomorpha*. **B)** Árbol filogenético que muestra los diferentes grupos de teleósteos y su relación evolutiva. Se indica la posición de las diferentes especies consideradas para el análisis, los diferentes parálogos (siguiendo un código de colores) presentes en cada especie y las pérdidas evolutivas de algunos de estos parálogos. Leyenda: TWGD: Duplicación genómica de teleósteos.

(Página anterior) **Figura Suplementaria 5 – Análisis de la estructura del dominio de activación transcripcional de Yap1b.**

Mapas de calor que representan el grado de similitud de secuencias (expresado en % de identidad de aminoácidos) entre proteínas de la familia Yap en diferentes especies para la proteína completa **(A)**, para los dominios TEAD **(B)** y para los dominios de transactivación **(C)**. Mientras que el dominio TEAD de Yap1b está relativamente bien conservado con respecto al dominio equivalente de Yap1, el dominio de transactivación muestra una menor homología (% de identidad de aa del 65.5% y del 44.8%, respectivamente). Alineamientos de secuencia del dominio de transactivación completo de los diferentes miembros de la familia Yap entre varias especies. A destacar, la pérdida o sustitución de tirosinas (azul) y fenilalaninas (rojo) conservadas en las proteínas Yap1b.

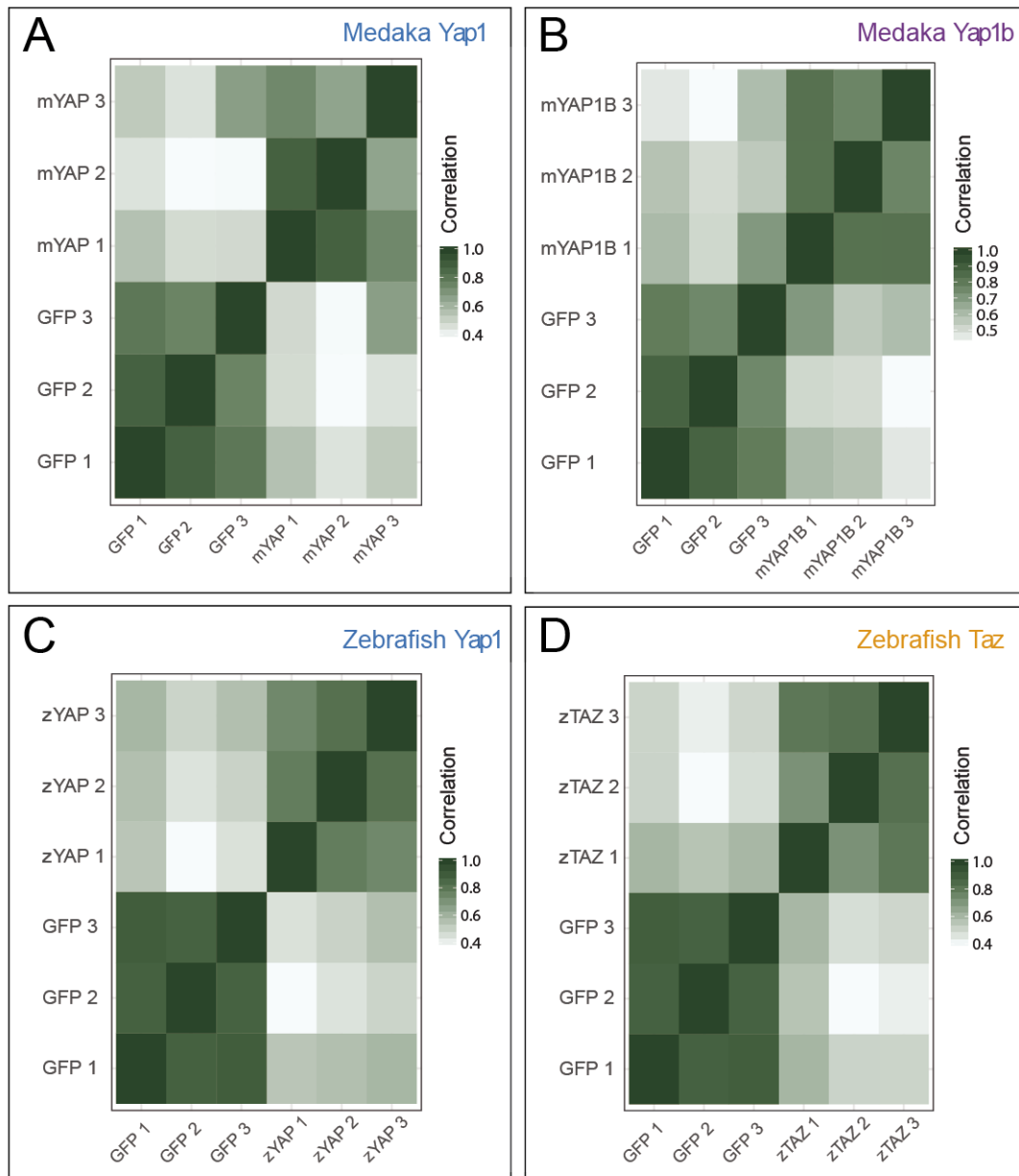
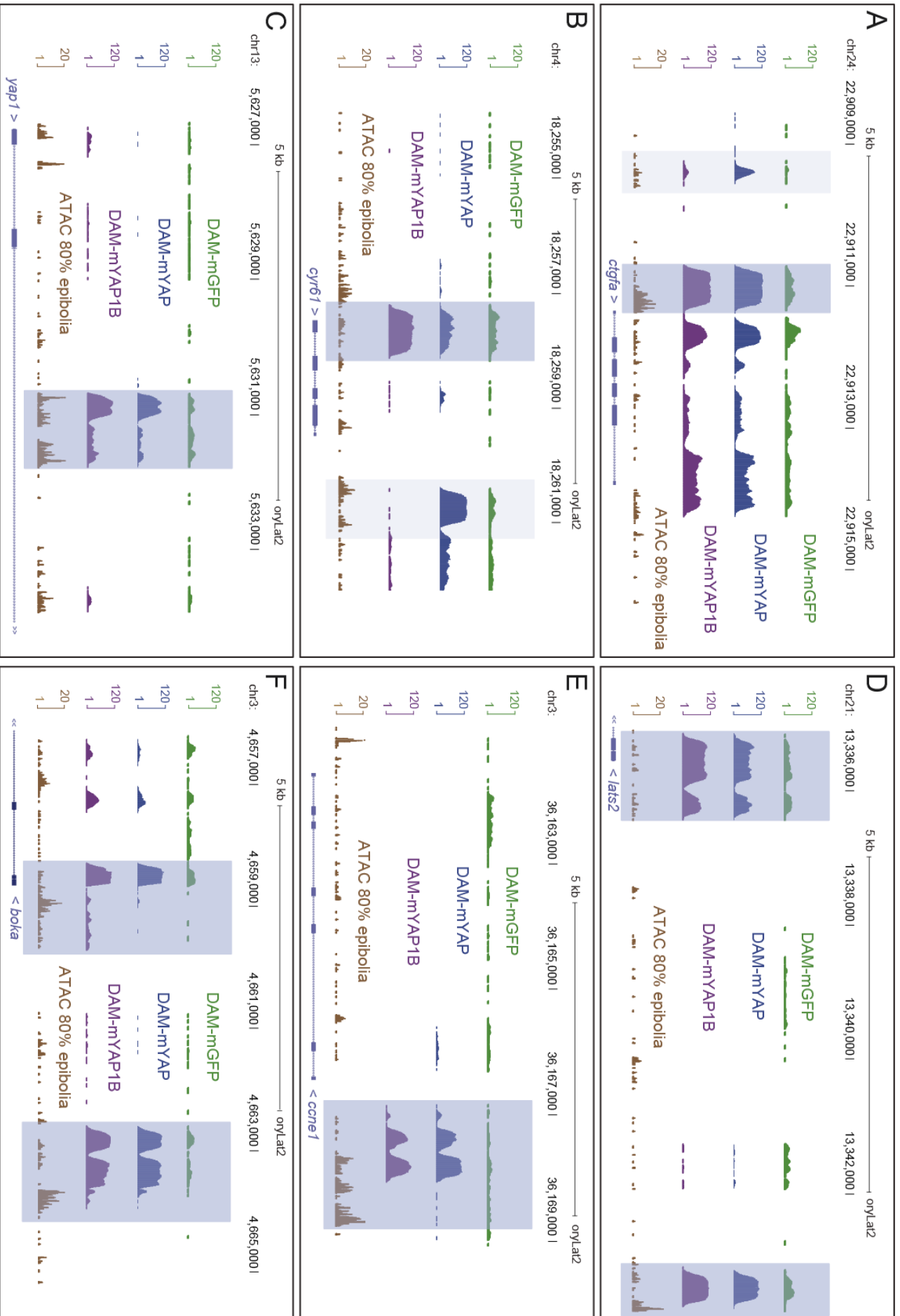


Figura Suplementaria 7 – Correlación entre réplicas en experimentos de iDamID-seq. Matriz que representa el coeficiente de correlación entre lecturas mapeadas entre réplicas (cada experimento se llevó a cabo por triplicado) y con respecto a sus correspondientes controles *DAM-GFP*, para cada muestra: *DAM-yap1* de medaka (**A**), *DAM-yap1b* de medaka (**B**), *DAM-yap1* de pez cebra (**C**) y *DAM-taz* de pez cebra (**D**).



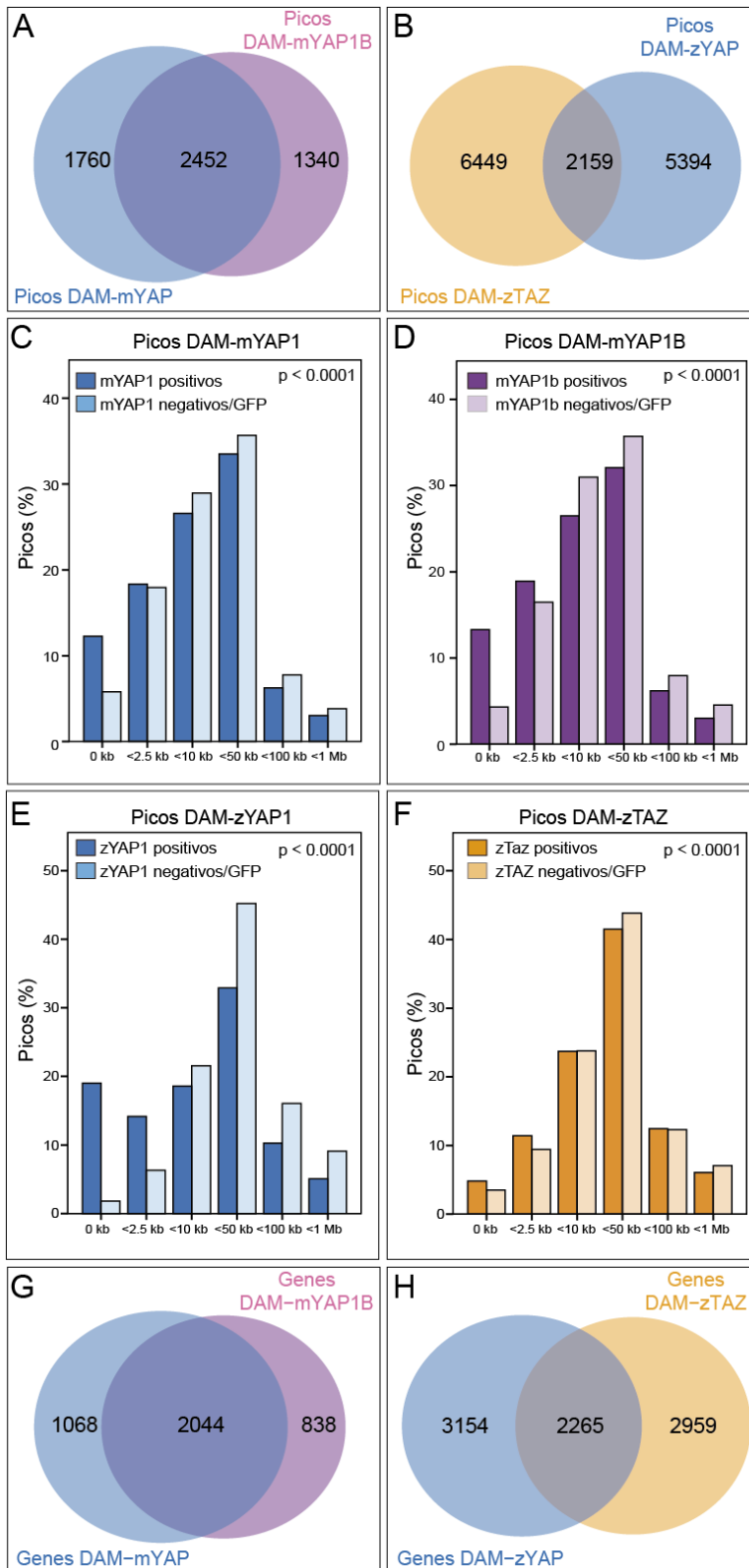


Figura Suplementaria 9 – Solapamiento de picos y distancia al TSS más cercano. Diagramas de Venn que muestran el grado de solapamiento entre picos *DAM-yap1* y *DAM-yap1b* de medaka (**A**) y entre picos *DAM-yap1* y *DAM-taz* de pez cebra (**B**). Se muestra la distribución de picos según su distancia al sitio de inicio de la transcripción (TSS) más cercano tanto para el conjunto de picos positivos como negativos (controles *DAM-GFP*) correspondientes a las muestras *DAM-yap1* y *DAM-yap1b* de medaka (**C, D**) y a las muestras *DAM-yap1* y *DAM-taz* de pez cebra (**E, F**). La distribución estadística de picos positivos frente a picos negativos fue determinada mediante un test de Chi-cuadrado (se indica el p-valor obtenido en cada caso). También se muestra el grado de solapamiento entre genes asociados a *DAM-yap1* y *DAM-yap1b* de medaka (**G**) y entre genes asociados a *DAM-yap1* y *DAM-taz* de pez cebra (**H**).

(Página anterior) **Figura Suplementaria 8 – Perfiles iDamID-seq adicionales de *yap1* y *yap1b* en medaka.** Perfiles genómicos de picos *DAM-yap1*, *DAM-yap1b* y *DAM-GFP* de embriones de medaka en estadio de 80% epibolia de varios genes diana de *yap1* como: *ctgfa* (**A**), *cyr61* (**B**), *yap1* (**C**), *lats2* (**D**), *ccne1* (**E**) y *boka* (**F**). Las señales DamID obtenidas también fueron comparadas con respecto a marcas de ATAC-seq. Algunos de los picos metilados con respecto al control que coinciden con regiones cis-reguladoras o intrónicas aparecen sombreados en azul.

A Motivos destacados de picos zYap1					B Motivos destacados de picos zTaz				
Motivo identificado	Motivo más parecido	Nombre	E-valor	q-valor	Motivo identificado	Motivo más parecido	Nombre	E-valor	q-valor
TAAAIA	IAAAIAAACA	FOX	$1.9 \cdot 10^{-373}$	$5.08 \cdot 10^{-2}$	CCTAGGGA	ATCCC GGGAA	EBF1	$3.9 \cdot 10^{-379}$	$9.29 \cdot 10^{-2}$
CATTAC	cATICC	TEAD	$7.4 \cdot 10^{-147}$	$2.84 \cdot 10^{-1}$	GGTCATGA	GGTCATGACC	RXRA	$1.3 \cdot 10^{-144}$	$1.30 \cdot 10^{-3}$
ACAAAG	ACAAATA	SOX	$1.1 \cdot 10^{-145}$	$3.18 \cdot 10^{-2}$	CCAGGC	CCAGGC	TFAP2	$5.5 \cdot 10^{-122}$	$1.57 \cdot 10^{-2}$
CGCGGC	CGCGC	ZFP161	$7.6 \cdot 10^{-59}$	$1.04 \cdot 10^{-2}$	AACATATG	AACATATG	NEUROG2 BHLHE22	$2.7 \cdot 10^{-98}$	$2.45 \cdot 10^{-2}$
AGGTTC	AGGTTCGTAAACA	ZNF435	$7.4 \cdot 10^{-53}$	$9.67 \cdot 10^{-2}$	ATCATT	ATCATTCAITC	ZNF24	$9.6 \cdot 10^{-66}$	$5.18 \cdot 10^{-3}$
GTCA CA	TcAGCA	MABF	$9.6 \cdot 10^{-47}$	$2.43 \cdot 10^{-2}$	GTCTAGAC	GTCTAGAC	SMAD3	$1.7 \cdot 10^{-52}$	$8.40 \cdot 10^{-4}$
AAAGCGG	ATGGCG	YY2	$5 \cdot 10^{-37}$	$2.58 \cdot 10^{-2}$	CAAGGA	CAGCAG	ZIC	$3.3 \cdot 10^{-45}$	$7.44 \cdot 10^{-2}$
ACATGTCC	ACATGTCAATGT	TP63	$7.2 \cdot 10^{-30}$	$9.72 \cdot 10^{-2}$	AAGCAAC	TGCCAAC	HIC1	$7.4 \cdot 10^{-31}$	$3.58 \cdot 10^{-2}$
TATCAA	TATGCAAAIT	POU	$3.6 \cdot 10^{-26}$	$7.39 \cdot 10^{-2}$	CCCAGC	CCCAGC	EGR1	$3.3 \cdot 10^{-24}$	$8.67 \cdot 10^{-4}$
ATTGCTIG	TTAATIG	HMX2	$5.3 \cdot 10^{-16}$	$3.10 \cdot 10^{-2}$	*****	*****			
					CATTCCCTC	cATICC	TEAD	$9.5 \cdot 10^{-3}$	$2.10 \cdot 10^{-1}$

Figura Suplementaria 10 – Identificación de motivos asociados a picos de iDamID-seq en pez cebra. Lista de motivos significativamente asociados a picos *DAM-yap1* (A) y *DAM-taz* (B) en pez cebra, identificados por DREME y reconocidos empleando la herramienta TomTom. Se muestran los E-valores de DREME y los q-valores de TomTom. El motivo TEAD está resaltado en rojo. TEAD es el vigesimosegundo motivo más representativo asociado a Taz (no se muestran los motivos del puesto 11 al 21 en la figura B).

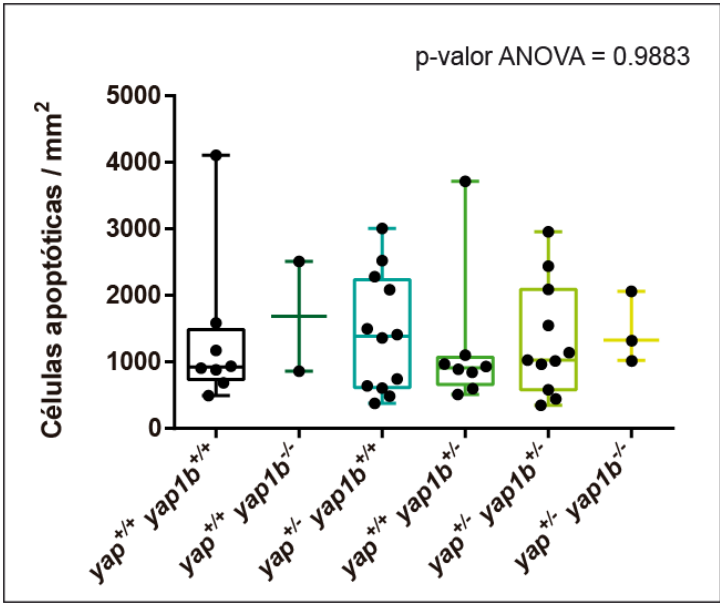


Figura Suplementaria 11 – Niveles de apoptosis en embriones de medaka con al menos una copia WT de *yap1*. La cuantificación de células positivas en caspasa-3 por área no muestra diferencias significativas entre genotipos mientras el embrión tenga al menos una copia silvestre de *yap1*. Se llevó a cabo un análisis ANOVA (One-way ANOVA) seguido de un test de Fisher para determinar el grado de significación entre genotipos concretos.

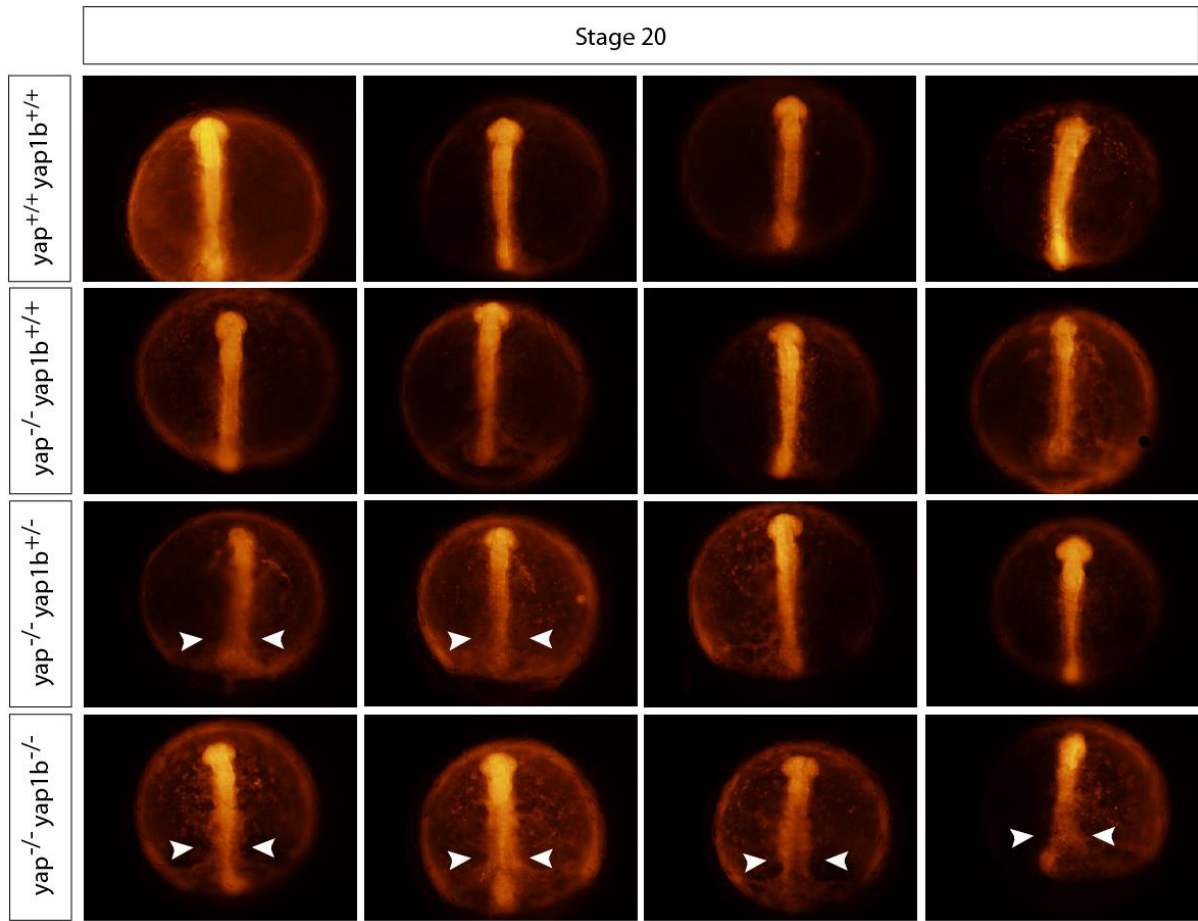


Figura Suplementaria 12 – Análisis fenotípico general de mutantes *yap1/yap1b*. Embriones mutantes yap1^{-/-} yap1b^{-/-}, yap^{-/-} yap1b^{+/-}, yap1^{-/-} yap1b^{+/+} y yap1^{+/+} yap1b^{+/+} en estadio 20, inyectados con un marcador de membrana. Obsérvese que, cuando no hay copias WT del gen *yap* y falta al menos una copia de *yap1b*, se aprecia cierta descondesación en la región posterior del embrión (marcado con flechas blancas). Este fenómeno resulta más evidente cuando faltan las dos copias de ambos genes.

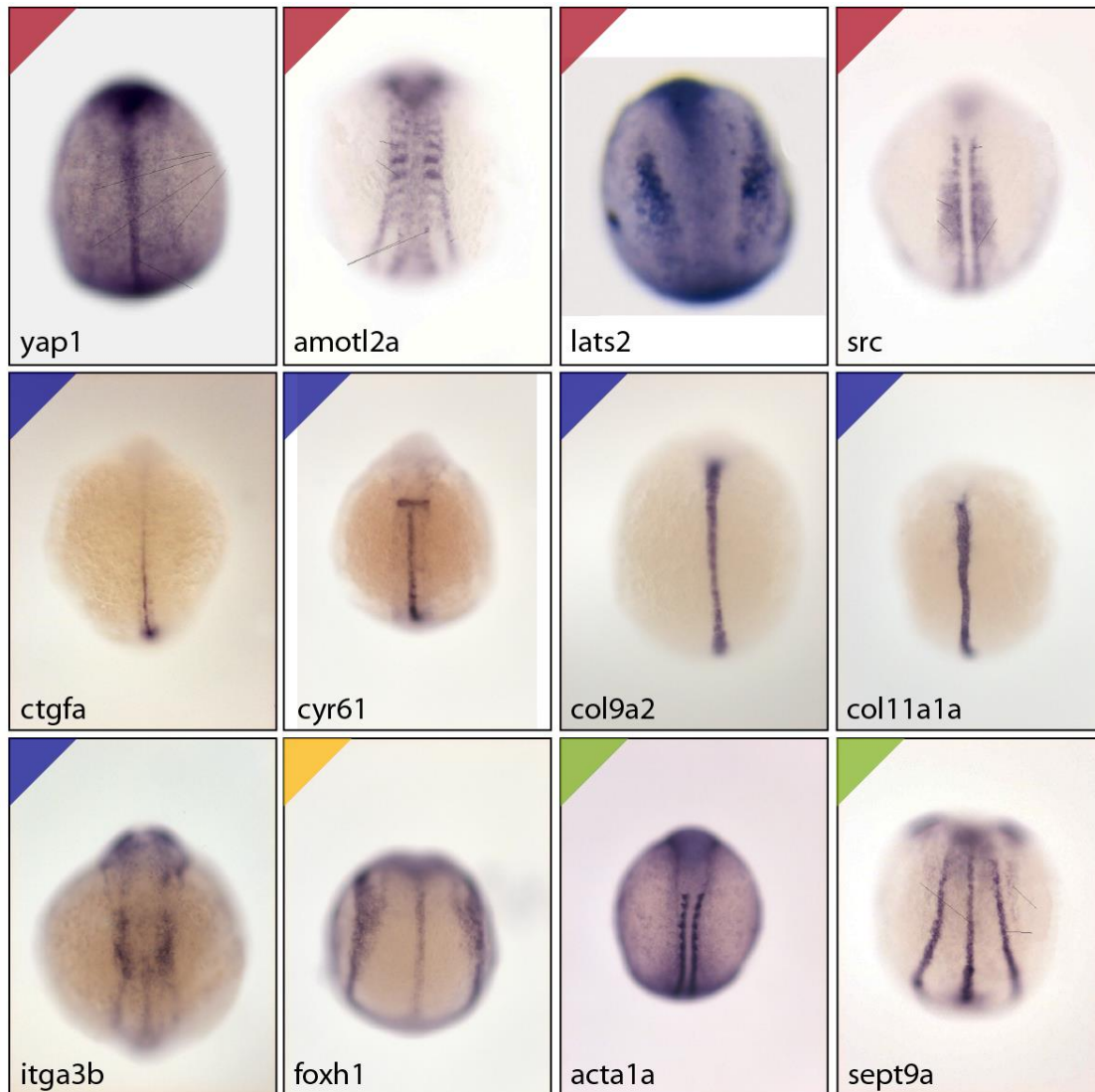


Figura Suplementaria 13 – Patrón de expresión en pez cebra de genes regulados por Yap1. Experimentos de ISH sobre genes asociados a Yap y a sus mecanismos de regulación (*yap1*, *amotl2a*, *lats2*, *src*; en rojo), de genes involucrados en procesos de adhesión (*ctgfa*, *cyr61*, *col9a2*, *col11a1a*, *itga3b*; en azul), de genes que codifican para diversos factores de transcripción (*foxh1*; en amarillo) y de genes implicados en la reorganización del citoesqueleto (*acta1a* y *sept9a*; en verde) en etapas posteriores a la gastrulación de pez cebra indica que muchos de ellos, cuya transcripción está directamente regulada por *yap1*, se expresan en la línea media o en tejidos adyacentes a ella.

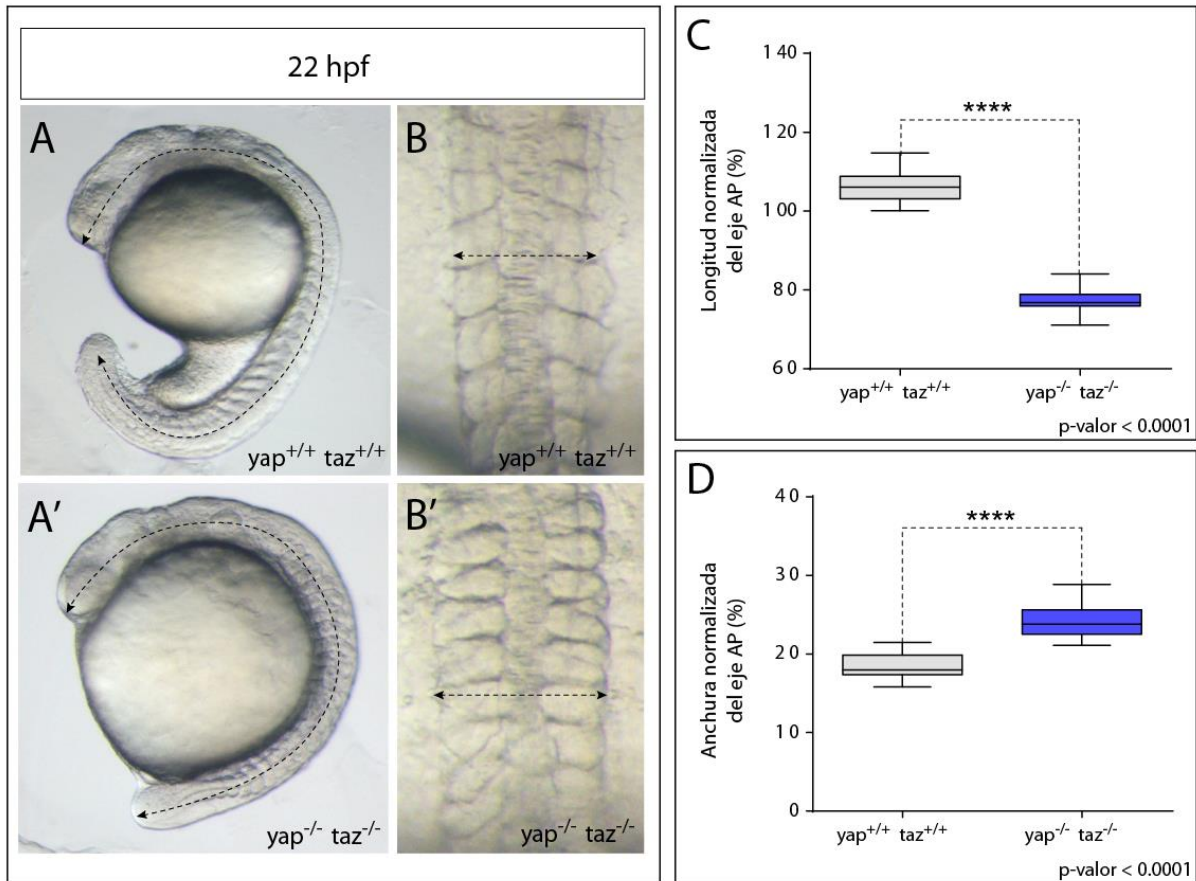


Figura Suplementaria 14 - Longitud y anchura de embriones *yap*^{-/-} *taz*^{-/-} de pez cebra. Los embriones WT de pez cebra presentan un eje anteroposterior más largo (**A**) y más estrecho (**B**) que los embriones *yap*^{-/-} (**A'**, **B'**). La normalización de la longitud del eje AP con respecto a la circunferencia del vitelo (**C**) y la del eje lateral con respecto al diámetro del vitelo (**D**) certifica cuantitativamente las diferencias fenotípicas entre ambos embriones. Se llevó a cabo un test de Fisher para determinar el grado de significación estadística de estas diferencias anatómicas.

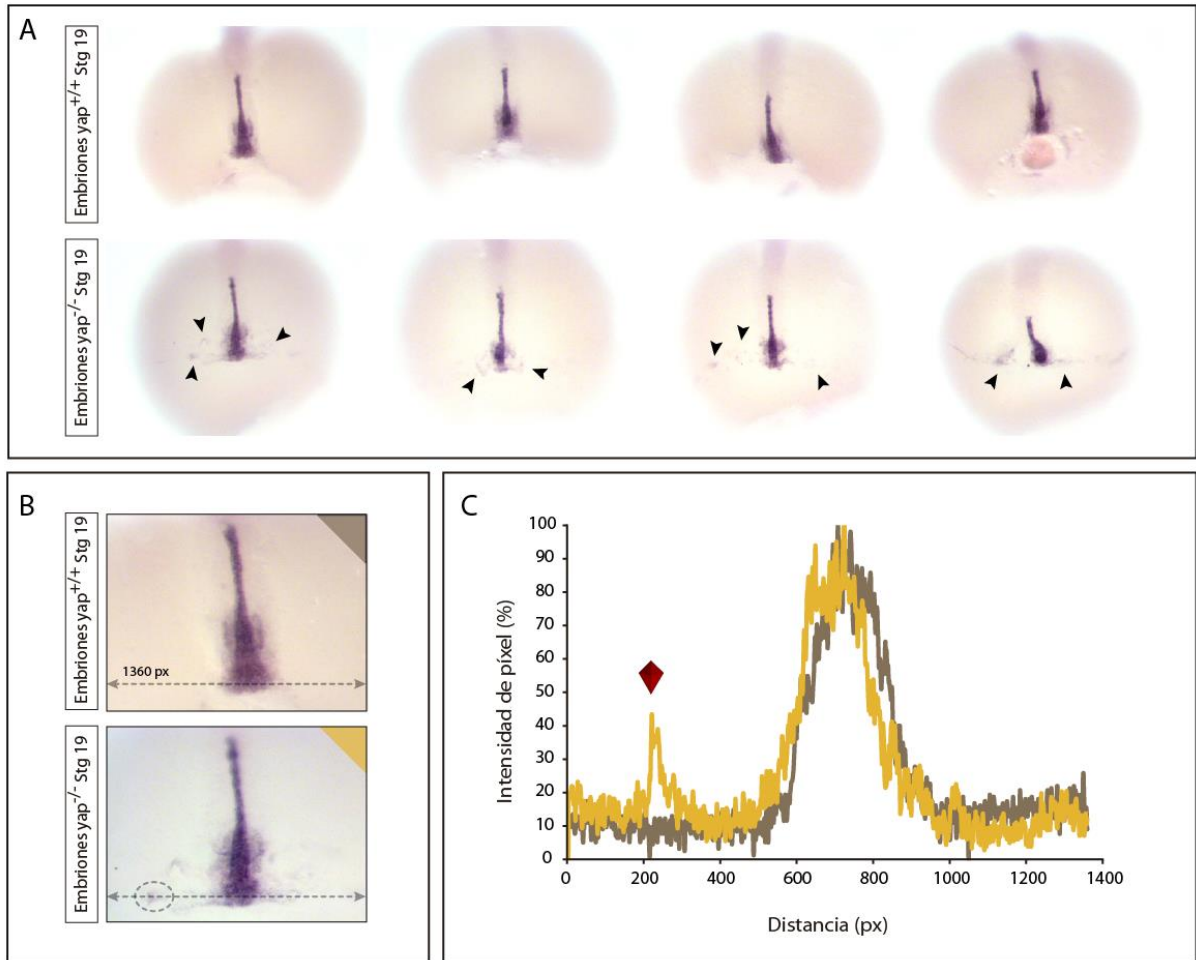


Figura Suplementaria 15 – Patrón de expresión de *ntl* en embriones *yap1*^{+/+} y *yap1*^{-/-} en embriones de medaka de estadio 19. **A)** El patrón de expresión de *ntl* está más definido en la parte posterior de la línea media de embriones *yap1*^{+/+}, mientras que en los embriones mutantes *yap1*^{-/-} se observan acumulaciones de células mesodérmicas a ambos lados de la línea media que no parecen haber migrado correctamente hacia ella (flechas negras). **B)** Se midió la intensidad de píxel de una sección transversal de 1360 px de longitud tomando una imagen de un embrión *yap1*^{+/+} y otra de un embrión *yap1*^{-/-}, de tal forma que ésta atravesase una región con células aisladas (circunferencia discontinua negra). **C)** Gráfica en la que aparece representada la intensidad relativa de cada píxel recogido en las secciones trazadas en el embrión WT (marrón) y mutante (naranja). Se aprecia con claridad la presencia de un pico en la región izquierda de la línea media (marcada con una señal roja) que corresponde al grupo de células aisladas indicado en la figura B.

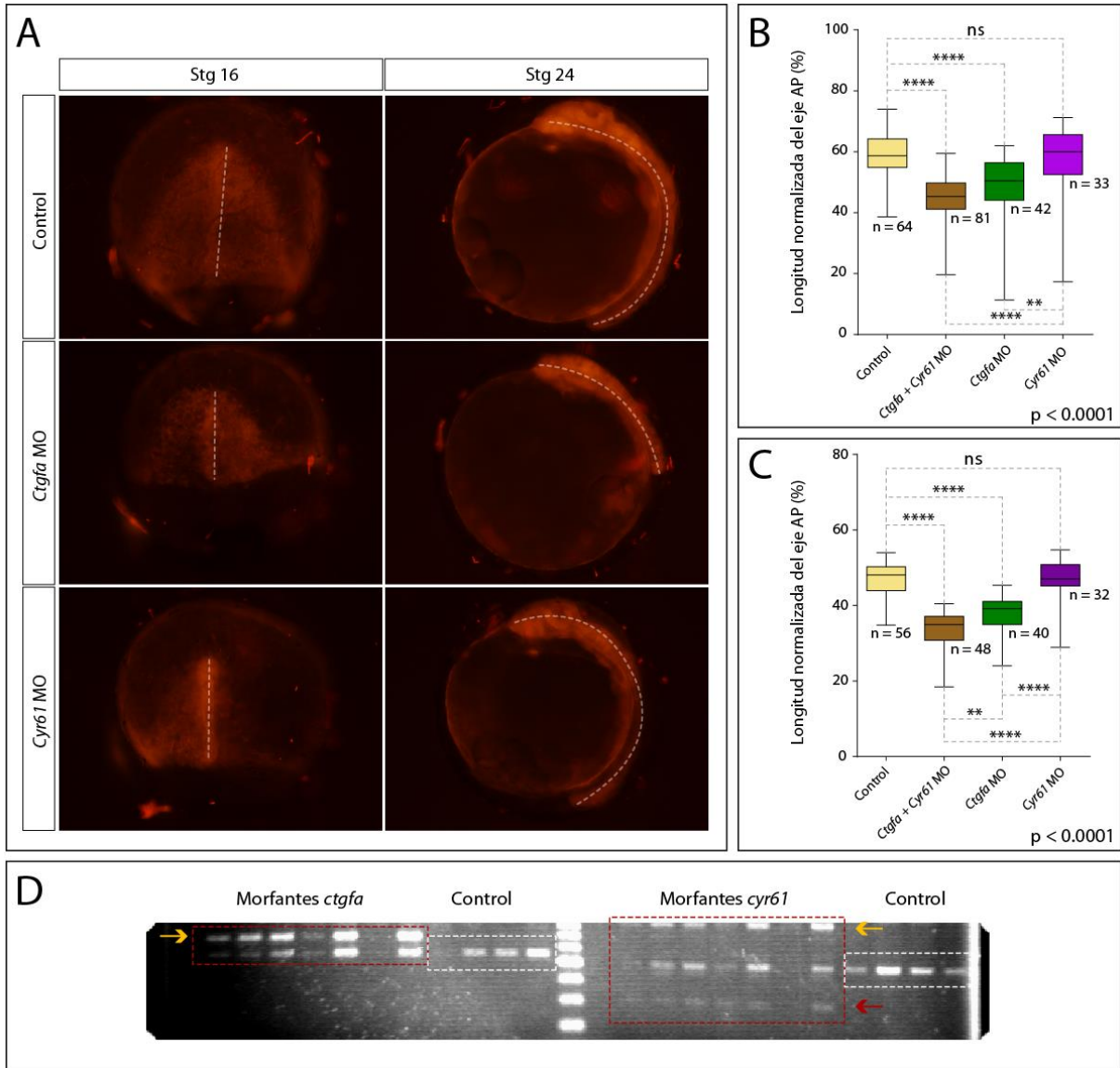


Figura Suplementaria 16 – Inyección por separado de morfolinos contra *ctgfa* y *cyr61* en medaka. A) La inyección de morfolinos contra el gen *ctgfa* en embriones de medaka provoca el acortamiento de su eje anteroposterior (línea discontinua blanca), hecho que ya es palpable desde el estadio 16 de desarrollo y que se mantiene incluso más allá del estadio 24. Por el contrario, no parece que la inyección de morfolinos contra el gen *cyr61* provoque efecto alguno en el embrión. Ambos morfolinos fueron coinyectados junto con el marcador de membrana LynTdtomato. **B)** Gráfica comparativa del grado de acortamiento del eje anteroposterior (normalizado con respecto al diámetro del vitelo) en embriones morfantes de *ctgfa*, morfantes de *cyr61*, morfantes dobles de ambos genes y embriones control, inyectados únicamente con el marcador de membrana, de estadio 16. **C)** Gráfica comparativa del grado de acortamiento del eje anteroposterior (normalizado con respecto a la circunferencia del vitelo) en embriones morfantes de *ctgfa*, morfantes de *cyr61*, morfantes dobles de ambos genes y embriones control, inyectados únicamente con el marcador de membrana, de estadio 24. Se llevó a cabo un análisis ANOVA (One-way ANOVA) y un test de Fisher para determinar el grado de significación estadística entre muestras. En las gráficas B y C se incluye el p-valor de ambos análisis ANOVA. **D)** Resultado de una PCR específica para analizar el funcionamiento de los morfolinos inyectados contra *ctgfa* (a la izquierda del marcador de peso molecular) y de los morfolinos inyectados contra *cyr61* (a la derecha del marcador de peso molecular). Se muestran los resultados de la PCR sobre seis embriones morfantes (resaltados en rojo) y cuatro embriones control (resaltados en blanco). En el caso de los morfantes de *ctgfa*, además de apreciarse la banda control esperada de 544 pb, se observa una banda de unos 670 pb (marcador amarillo), que corresponderían a un caso de retención del intrón 1 (vea *Materiales y métodos*). En el caso de los morfantes de *cyr61*, además de la banda control de 520 pb, se aprecia una banda de unos 800 pb (que correspondería a la retención del intrón 2, de 317 pb de longitud) y una banda más pequeña, de alrededor de 300 pb (que correspondería a la pérdida del exón 2 del gen, de 232 pb de longitud) (vea *Materiales y métodos*).

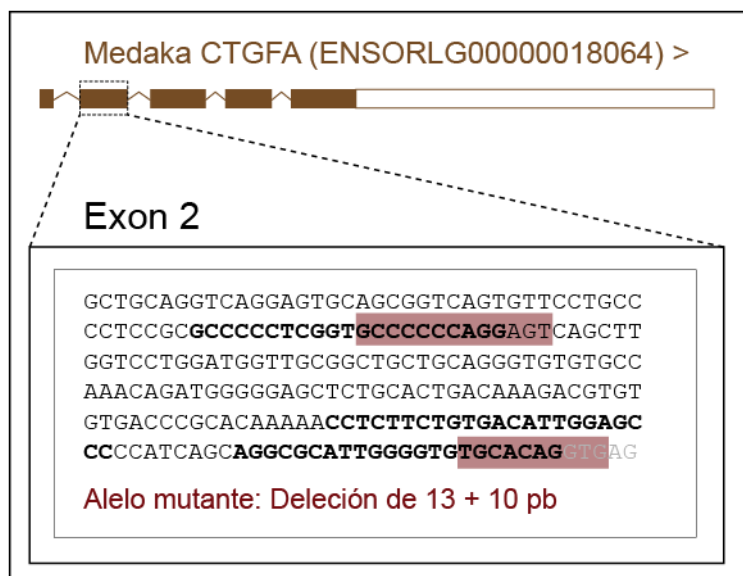


Figura Suplementaria 17 – Generación de alelos mutantes para el gen *ctgfa* de medaka. Se utilizó la tecnología CRISPR-Cas9 para establecer líneas mutantes del gen *ctgfa* en pez medaka. Se inyectaron cuatro moléculas de sgRNA, tres de las cuales tienen su diana en el exón 2 del gen (en negrita). Por el momento, se ha logrado identificar y extender un alelo mutante, consistente en una delección doble de 13 y 10 pb (sombreado en rojo), que genera una proteína truncada. Parte de la delección de 10 pb se produce en el intrón 2 del gen (en gris).

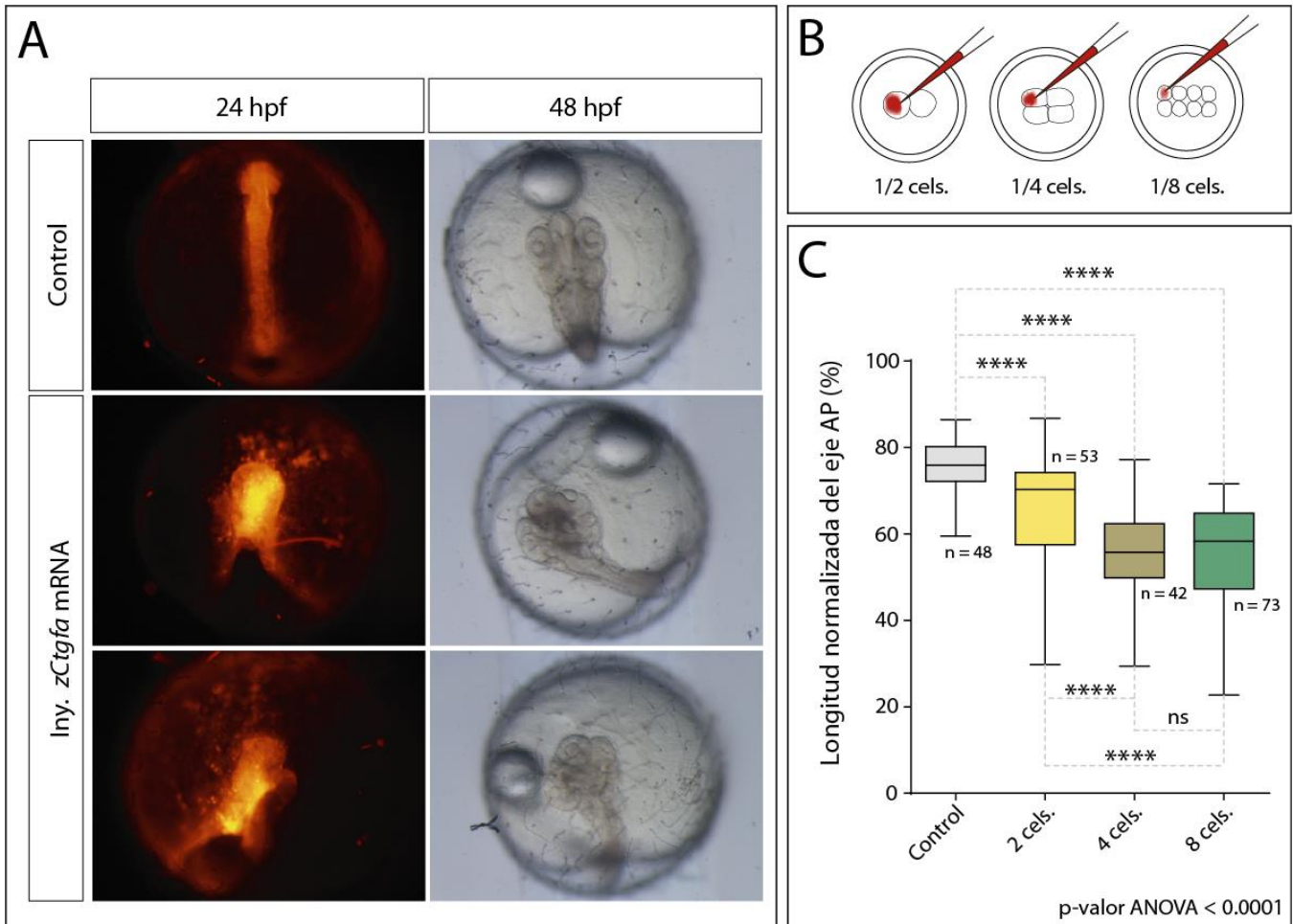


Figura Suplementaria 18 – Inyección asimétrica de *zCtgfa* en embriones de medaka para establecer gradientes. **A)** La inyección de moléculas de mRNA de pez cebra exclusivamente en una sola célula de embriones de medaka de estadios de 2, 4 y 8 células genera embriones con un eje anteroposterior anormalmente corto y con un frente de epibolia cóncavo, en forma de “V”. Estas anomalías se traducen posteriormente en embriones con importantes malformaciones. **B)** Esquema indicativo que muestra de manera sencilla cómo se generaron los tres gradientes de *zCtgfa*. **C)** El acortamiento del eje anteroposterior es muy significativo al establecer estos gradientes en estadio de dos células y especialmente en embriones en estadio de 4-8 células.

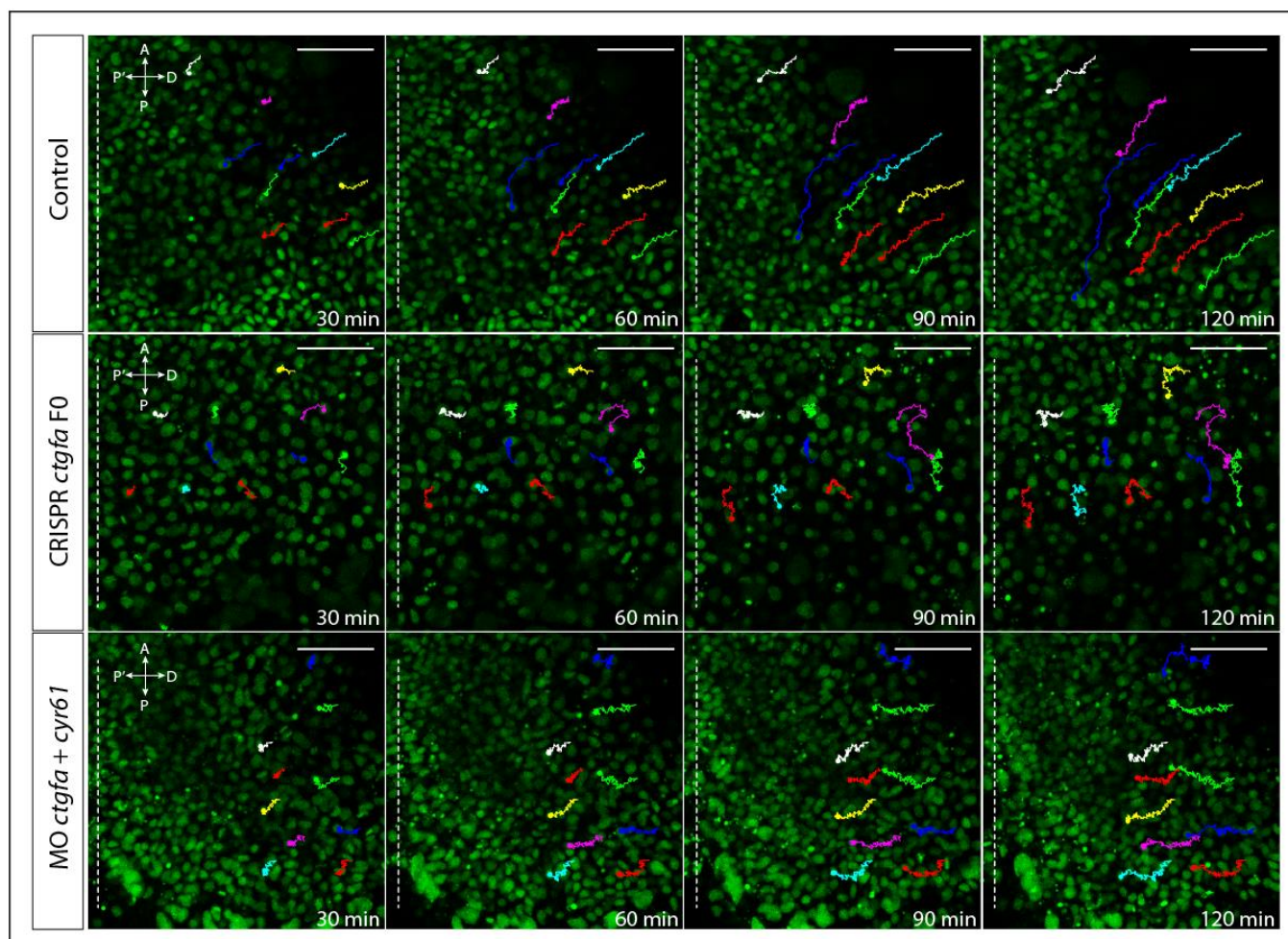


Figura Suplementaria 19 – Time lapse de la parte posterior de embriones de medaka durante la gastrulación. Se representan las trayectorias de 10 células de la región posterior de un embrión WT, un embrión CRISPante y un embrión morfante de estadio 16 a lo largo de un periodo de 120 minutos, en intervalos de 30 minutos. Leyenda: Línea discontinua: línea media; A: región anterior; P: región posterior; P': región proximal a la línea media; D: región distal a la línea media. Escala: 50 μ m.

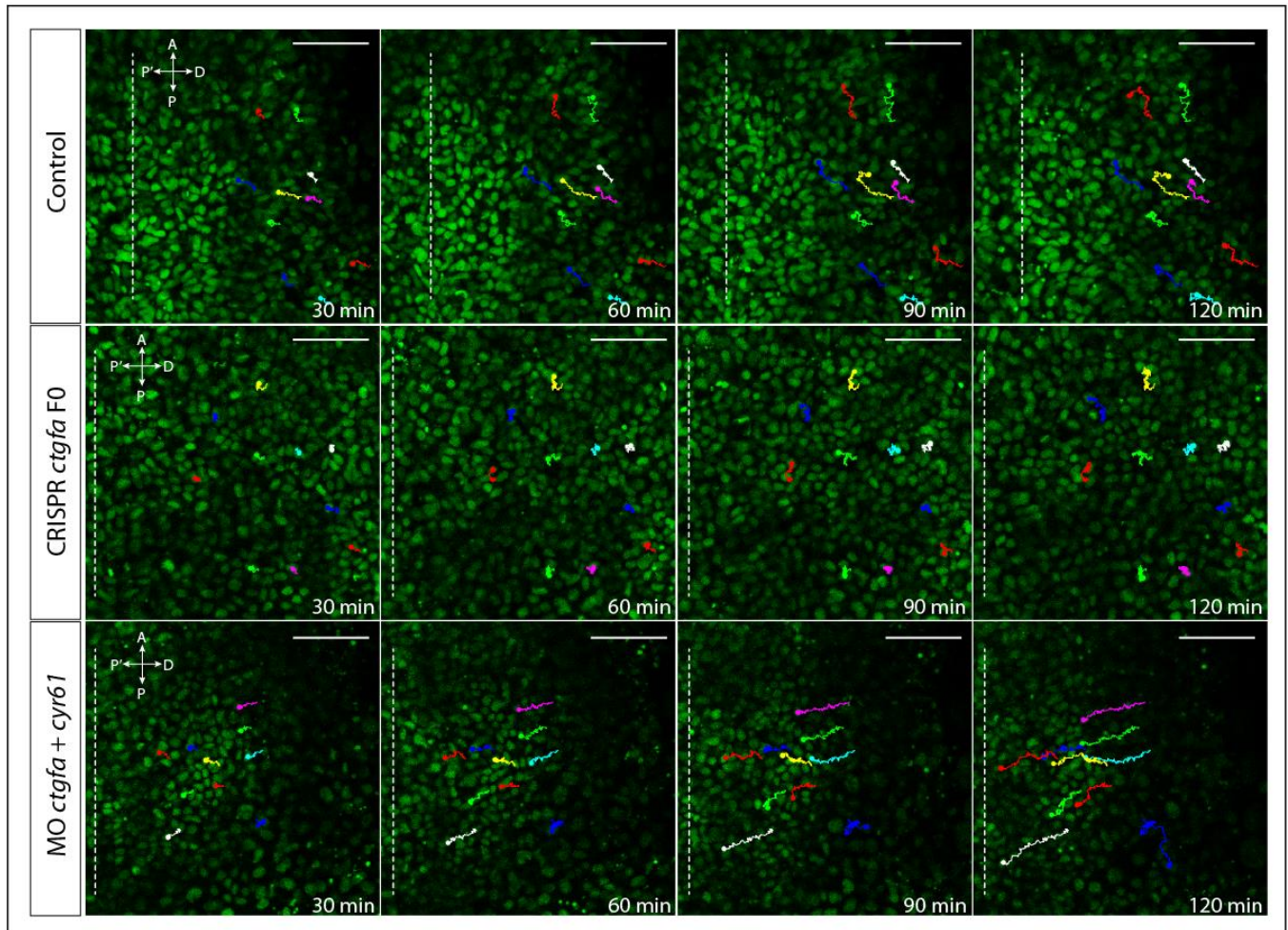


Figura Suplementaria 20 – Time lapse de la parte anterior de embriones de medaka durante la gastrulación. Se representan las trayectorias de 10 células de la región anterior de un embrión WT, un embrión CRISPRante y un embrión morfante de estadio 16 a lo largo de un periodo de 120 minutos, en intervalos de 30 minutos. Leyenda: Línea discontinua: línea media; A: región anterior; P: región posterior; P': región proximal a la línea media; D: región distal a la línea media. Escala: 50 μ m.

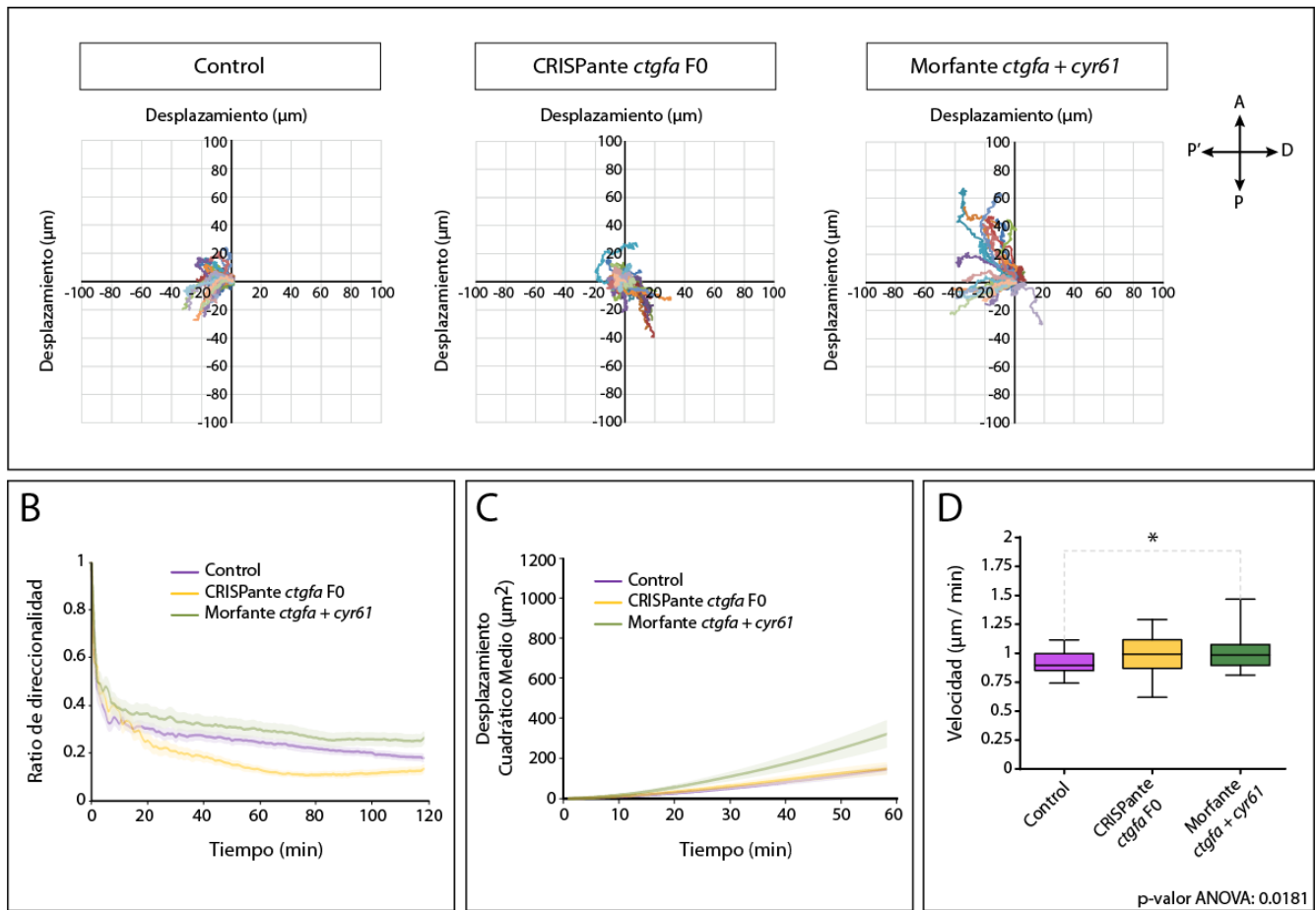


Figura Suplementaria 21 – Rastreo de trayectorias celulares en la región anterior de embriones de medaka durante la gastrulación. **A)** Se determinan las trayectorias de una muestra de 30 células correspondientes a tres embriones WT, tres CRISPantes de *ctgfa* y tres morfantes de *ctgfa* y *cyr61* de estadio 16 a lo largo de un lapso de tiempo de 2 horas, tomadas de películas grabadas in vivo en un microscopio confocal. Las células WT migran hacia la línea media, pero mucho menos en comparación con las de la región posterior, pues allí el proceso de convergencia-extensión aún se encuentra en sus primeras etapas. Las células CRISPantes no parecen migrar de manera normal a la línea media, mientras que las células morfantes, aunque sí lo hacen, tienden a desplazarse una mayor distancia y hacia el polo animal, lo que sugiere un retraso en el proceso de convergencia-extensión. **B)** Ratio de direccionalidad de células control (púrpura), CRISPantes (naranja) y morfantes (verde). La trayectoria de las células CRISPantes parece más errática que en las otras dos condiciones. **C)** Desplazamiento cuadrático medio (MSD) de células control, CRISPantes y morfantes durante los primeros 60 minutos. Las células morfantes presentan un MSD superior, pero en cualquier caso, estas medidas son muy inferiores a las obtenidas en células de la parte posterior del embrión. **D)** Velocidad media de células control, CRISPantes y morfantes. La velocidad prácticamente no varía entre condiciones, aunque las células morfantes son ligeramente más veloces que las células control. En las gráficas B y C, el trazo grueso de cada condición determina el valor promedio, mientras que las regiones sombreadas adyacentes representan su error estándar. Para el análisis de velocidad celular, se llevó a cabo un análisis One-way ANOVA para determinar el grado de significación de las diferencias observadas de velocidad entre las tres condiciones. El grado de significación comparativo uno a uno entre condiciones se efectuó mediante un test de Fisher. Leyenda: A: región anterior; P: región posterior; P: extremo proximal a la línea media; D: extremo distal a la línea media.

Nº	Nombre	Secuencia (5' - 3')	Sección
1	mYapKOsg#3_TAGG	TAGGCACGTCGTCGGGAATCTCGT	3
2	mYapKOsg#3_AAAC	AAACACGAGATCCCCGACGACGTG	3
3	mYapKOsg#4_TAGG	TAGGATCTCGTAGGACGACTGC	3
4	mYapKOsg#4_AAAC	AAACGCAGTCGCTCTACGAGAT	3
5	mYapKOsg#5_TAGG	TAGGGACGAGTGTGCGCGTACG	3
6	mYapKOsg#5_AAAC	AAACCGTACGCGCACACTCGTC	3
7	CRISPR mYAP Fw	CAGCCCACAAAACATACAGTTTA	3
8	CRISPR mYAP Rv	AAGATCTGCTGGGAGAACTCTTT	3
9	sgRNA mYAP1B #1	taatacgactcactataGGGGGGAGTGGCACATCGTCgttttagagctagaa	3
10	sgRNA mYAP1B #2	taatacgactcactataGGCGGGGAGCGCAGCGGGGGgttttagagctagaa	3
11	CRISPR mYAP1B Fw	TTTAAAGAAACCTGCAGAGCAAC	3
12	CRISPR mYAP1B Rv	AGCAACATGTCCAATTA AAAACC	3
13	sgRNA mCtgfa #1	taatacgactcactataGGGCCCCCTCGGTGCCCCCgttttagagctagaa	3
14	sgRNA mCtgfa #2	taatacgactcactataGGGCGCATTGGGGTGTGCACgttttagagctagaa	3
15	sgRNA mCtgfa #3	taatacgactcactataGGGCTCCAATGTCACAGAAGgttttagagctagaa	3
16	sgRNA mCtgfa #4	taatacgactcactataGGGTAAATCAAGCAGAGCAGgttttagagctagaa	3
17	CRISPR mCtgfa Fw	TCCACTTGATCCGAAGTTCC	3
18	CRISPR mCtgfa Rv	GCCTCCAAACACACAAGAGG	3
19	OL-YAP Fw-SpeI	AATTTactagtATGGATCCGAGCCAGCACAAACCTC	4
20	OL-YAP Rv-XbaI	AATTTtctagaTTATAACCATGTGAGGAAGCTCTC	4
21	OL-YAP1B Fw-SpeI	AATTTactagtATGGACGCGCACCAGGAGCGGACG	4
22	OL-YAP1B Rv-NheI	AATTTgctagcCTATAGCCAGGTGAGCAGATTGTC	4
23	DR-YAP Fw-SpeI	AATTTactagtATGGATCCGAACCAGCACAAAC	4
24	DR-YAP Rv-NheI	AATTTgctagcCTATAGCCAGGTAGAAAGTTC	4
25	DR-TAZ Fw-SpeI	AATTTactagtATGAGCGGTAATCCTCTCCAGC	4
26	DR-TAZ Rv-XbaI	AATTTtctagaTTAGAGCCAGGTGAGGAAGGGCTC	4
27	AdRt	CTAATACGACTCACTATAGGGCAGCGTGGTCGCGCCGAGGA	4
28	AdRb	TCCTCGGCCG	4
29	AdR_PCRprimer	G*G*T*C*G*C*GGCCGAGGATC	4
30	ISH ntl MDK Fw	AGAGTTAACCAACGAGATGATCG	6
31	ISH ntl MDK Rv	CAACCTCCAAGTTGGGAGTATC	6
32	ISH yap MDK Fw	ACTCCAGATGACTTCCTCAACAG	6
33	ISH yap MDK Rv	AGCCTTGAAGACACAGACACAAT	6
34	ISH yap1b MDK Fw	CCCTGTTCACTAGGGTGATGA	6
35	ISH yap1b MDK Rv	TACATCTCTCCTCTGGAGTCAC	6
36	qPCR zEF1A Fw	TCCACCGGTACCTGATCTACA	7
37	qPCR zEF1A Rv	CAACACCCAGGCGTACTTGA	7
38	qPCR zYAP Fw	ATGAACCCAGCCTCAGGTCC	7
39	qPCR zYAP Rv	TCCAGCCAAGAGGTGGTTTT	7
40	qPCR zTAZ Fw	TTCCTCAACAGTGGGCCGTA	7
41	qPCR zTAZ Rv	AGTGGTCGGGATGCTGTAAC	7
42	qPCR mEF1A Fw	AAACCCAGAAACACCGAAACAT	7
43	qPCR mEF1A Rv	CCTCCGCACTTGTAGATCAG	7
44	qPCR mYAP Fw	GTTGAACCCAGGCTACCCCTC	7
... (Continúa en la página siguiente)			

Nº	Nombre	Secuencia (5' - 3')	Sección
45	qPCR mYAP Rv	AGGTCCACTAGCTGGGCTTA	7
46	qPCR mYAP1B Fw	GTCAACCCGGAGTCAGGTCC	7
47	qPCR mYAP1B Rv	AGTACATCTCTCCCTCTGGAGTC	7
48	ScMOs_mCtgF Fw	GTTTGCTAACAAGGAGAAGGTCA	8
49	ScMOs_mCtgF Rv	CTCCTCTCTGAAAGCAGACAAAG	8
50	ScMOs_mCyr61 Fw	TGCTTCATTCTCTCGTTTCTTC	8
51	ScMOs_mCyr61 Rv	GGATATCTGTCTCCTTGCCATCT	8
52	EcoRI-zCTGF CDS Fw	AATTTgaattcATGTTTCTGGAATGACTCAAAGTG	8
53	XbaI-zCTGF CDS Rv	AATTTtctagaTCACGCCATGTCGCCAACCATCTTC	8

Tabla Suplementaria 1 – Lista de oligonucleótidos utilizados en este proyecto de tesis. Aparece representado el número y nombre identificativo de cada uno, su secuencia y la sección de *Materiales y métodos* en donde aparecen detalladas las razones de su uso.

Tabla Suplementaria 2: Lista de identificadores de genes y proteínas de la familia Yap de diferentes especies. También se indica el cromosoma al que pertenece cada parólogo y los diferentes cambios de sintenia observados en cada especie. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/TablaSup2.xlsx>

Tabla Suplementaria 3: Lista completa de picos iDamID-seq de *yap1* y *yap1b* de medaka, y de *yap1* y *taz* de pez cebra. Se representan las coordenadas de cada pico y las de su TSS más cercano, indicando también su identificador y su distancia exacta en pares de bases. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/TablaSup3.xlsx>

Tabla Suplementaria 4: Términos GO-Slim sobrerrepresentados del conjunto de genes asociados a *yap1* y *yap1b* en medaka. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/TablaSup4.xlsx>

Tabla Suplementaria 5: Datos de expresión de genes identificados en el RNA-seq. Se muestra el log₂ del nivel de expresión diferencial entre genes reprimidos (marcados en rojo) y activados en mutantes *yap1*^{-/-} (marcados en verde). Para cada gen diferencialmente expresado, se indica su identificador, su nombre y sus coordenadas. Sólo se muestran aquellos genes con un nivel

de expresión diferencial entre WT y mutantes estadísticamente significativo. Descárguela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/TablaSup5.xlsx>

ID	Gene name	ID	Gene name
ENSORLG00000018064	ctgfa	ENSORLG00000005834	apbb2b
ENSORLG00000014159	marcksl1b	ENSORLG00000015826	MPZL2 (1 of many)
ENSORLG00000002708	yap1	ENSORLG00000000698	CARMIL1 (1 of many)
ENSORLG00000004199	-	ENSORLG00000002876	elovl6l
ENSORLG00000002123	boka	ENSORLG00000001982	flna
ENSORLG00000018947	-	ENSORLG00000009834	fermt1
ENSORLG00000013339	XYLT1 (1 of many)	ENSORLG00000017692	egfra
ENSORLG00000008004	zgc:101810	ENSORLG00000004983	oncut3a
ENSORLG00000016970	zgc:158343	ENSORLG00000000682	SPHK1 (1 of many)
ENSORLG00000008993	ptk2bb	ENSORLG00000009563	-
ENSORLG00000009065	SHISA2	ENSORLG00000001801	-
ENSORLG00000009627	nt5c2l1	ENSORLG00000015829	tnfaip8l2b
ENSORLG00000002316	neurl1aa	ENSORLG00000016183	src
ENSORLG00000012314	rnd1a	ENSORLG00000007918	slit1a
ENSORLG00000012170	MAFG (1 of many)	ENSORLG00000015892	serpina1l
ENSORLG00000003196	sqlea	ENSORLG00000018633	-
ENSORLG00000010849	SYDE2	ENSORLG00000007736	gata2a
ENSORLG00000005664	col11a1a	ENSORLG00000004795	FZD8 (1 of many)
ENSORLG00000008347	trim45	ENSORLG00000015426	rbms2a
ENSORLG00000017569	marcksl1a	ENSORLG00000003829	ptger4b
ENSORLG00000003107	foxo1	ENSORLG00000015478	lrp1aa
ENSORLG00000015206	adamts15b	ENSORLG00000015529	pik3cb
ENSORLG00000000178	ADAchrMSL5	ENSORLG00000005951	DLC1
ENSORLG00000013304	bmp4	ENSORLG00000004948	-
ENSORLG00000016273	amotl2b	ENSORLG00000007331	foxh1
ENSORLG00000012100	ftr82	ENSORLG00000005252	-
ENSORLG00000001604	NTF4	ENSORLG00000008517	VCL (1 of many)
ENSORLG00000011536	pim3	ENSORLG00000016366	ptger1b
ENSORLG00000014908	PTPN14	ENSORLG00000004191	rapgef1a
ENSORLG00000015488	bmper	ENSORLG00000004493	bambia
ENSORLG00000015665	tp63	ENSORLG00000011645	mid1ip1a
ENSORLG00000003407	TIAM1	ENSORLG00000010424	-
ENSORLG00000009681	-	ENSORLG00000000886	pcdh19
ENSORLG00000009419	actn1	ENSORLG00000012987	tubb5
ENSORLG00000017671	fam129bb	ENSORLG00000009818	gpc4
ENSORLG00000015457	amotl2a	ENSORLG00000000140	mfsd6a
ENSORLG00000013186	lats2		

Tabla Suplementaria 6 – Lista completa de genes reprimidos en mutantes *yap1*^{-/-} de medaka asociados a picos iDamID-seq. Para cada uno de los 73 genes indicados, se muestran su identificador y su nombre.

Película Suplementaria 1: Time-lapse de 960 minutos de un embrión WT de medaka durante las últimas etapas de la gastrulación. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup1.xlsx>

Película Suplementaria 2: Time-lapse de 900 minutos de un embrión CRISPante de *ctgfa* de medaka durante las últimas etapas de la gastrulación. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup2.xlsx>

Película Suplementaria 3: Time-lapse de 840 minutos de un embrión morfante de *ctgfa* y *cyr61* de medaka durante las últimas etapas de la gastrulación. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup3.xlsx>

Película Suplementaria 4: Time-lapse detallado de 120 minutos de la región posterior de un embrión WT de medaka durante las últimas etapas de la gastrulación. Se muestran las trayectorias de 10 células. La línea media del embrión se sitúa en el lado izquierdo de la imagen y la región del polo vegetal, en la zona inferior. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup4.xlsx>

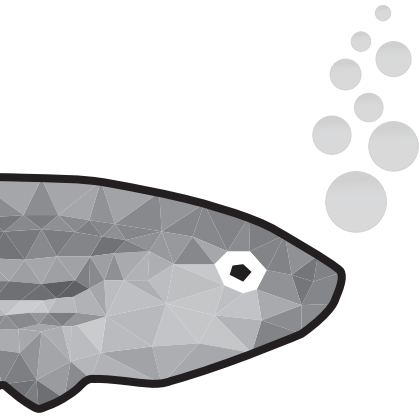
Película Suplementaria 5: Time-lapse detallado de 120 minutos de la región posterior de un embrión CRISPante de *ctgfa* de medaka durante las últimas etapas de la gastrulación. Se muestran las trayectorias de 10 células. La línea media del embrión se sitúa en el lado izquierdo de la imagen y la región del polo vegetal, en la zona inferior. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup5.xlsx>

Película Suplementaria 6: Time-lapse detallado de 120 minutos de la región posterior de un embrión morfante de *ctgfa* y *cyr61* de medaka durante las últimas etapas de la gastrulación. Se muestran las trayectorias de 10 células. La línea media del embrión se sitúa en el lado izquierdo de la imagen y la región del polo vegetal, en la zona inferior. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup6.xlsx>

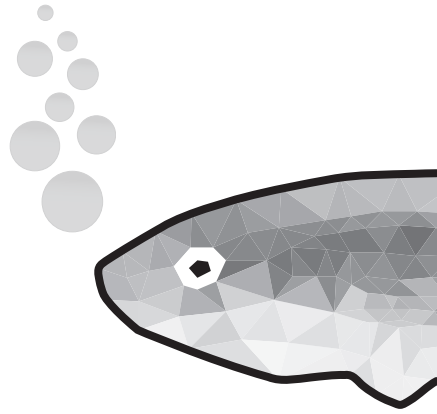
Película Suplementaria 7: Time-lapse detallado de 120 minutos de la región cefálica de un embrión WT de medaka durante las últimas etapas de la gastrulación. Se muestran las trayectorias de 10 células. La línea media del embrión se sitúa en el lado izquierdo de la imagen y la región del polo animal, en la zona superior. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup7.xlsx>

Película Suplementaria 8: Time-lapse detallado de 120 minutos de la región cefálica de un embrión CRISPR de *ctgfa* de medaka durante las últimas etapas de la gastrulación. Se muestran las trayectorias de 10 células. La línea media del embrión se sitúa en el lado izquierdo de la imagen y la región del polo animal, en la zona superior. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup8.xlsx>

Película Suplementaria 9: Time-lapse detallado de 120 minutos de la región cefálica de un embrión morfante de *ctgfa* y *cyr61* de medaka durante las últimas etapas de la gastrulación. Se muestran las trayectorias de 10 células. La línea media del embrión se sitúa en el lado izquierdo de la imagen y la región del polo animal, en la zona superior. Descárguesela a través del siguiente enlace: <http://193.147.188.155/jvazmar/Tesis/PeliculaSup9.xlsx>



Referencias bibliográficas



- Agarwala, S., Duquesne, S., Liu, K., Boehm, A., Grimm, L., Link, S., ... Lecaudey, V. (2015). Amotl2a interacts with the hippo effector yap1 and the Wnt/ β -catenin effector lef1 to control tissue size in zebrafish. *ELife*, 4(September 2015), 1–26. <http://doi.org/10.7554/eLife.08201>
- Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., ... Postlethwait, J. H. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science*, 282(5394), 1711–1714. <http://doi.org/10.1126/science.282.5394.1711>
- Aragona, M., Panciera, T., Manfrin, A., Giullitti, S., Michielin, F., Elvassore, N., ... Piccolo, S. (2013). A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. *Cell*, 154(5), 1047–59. <http://doi.org/10.1016/j.cell.2013.07.042>
- Aughey, G. N., & Southall, T. D. (2016). Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdisciplinary Reviews: Developmental Biology*, 5(1), 25–37. <http://doi.org/10.1002/wdev.205>
- Azzolin, L., Panciera, T., Soligo, S., Enzo, E., Bicciato, S., Dupont, S., ... Piccolo, S. (2014). YAP/TAZ incorporation in the β -catenin destruction complex orchestrates the Wnt response. *Cell*, 158(1), 157–170. <http://doi.org/10.1016/j.cell.2014.06.013>
- Bailey, T. L. (2011). DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, 27(12), 1653–1659. <http://doi.org/10.1093/bioinformatics/btr261>
- Bandura, J. L., & Edgar, B. A. (2008). Yorkie and Scalloped: partners in growth activation. *Developmental Cell*, 14(3), 315–316. <http://doi.org/10.1016/j.devcel.2008.02.010>
- Baumgartner, R., Poernbacher, I., Buser, N., Hafen, E., & Stocker, H. (2010). The WW Domain Protein Kibra Acts Upstream of Hippo in Drosophila. *Developmental Cell*, 18(2), 309–316. <http://doi.org/10.1016/j.devcel.2009.12.013>
- Benjamin, J. M., & Nelson, W. J. (2008). Bench to bedside and back again: Molecular mechanisms of α -catenin function and roles in tumorigenesis. *Seminars in Cancer Biology*, 18(1), 53–64. <http://doi.org/10.1016/j.semcancer.2007.08.003>
- Bensch, R., Song, S., Ronneberger, O., & Driever, W. (2013). Non-directional radial intercalation dominates deep cell behavior during zebrafish epiboly. *Biology Open*, 2(8), 845–854. <http://doi.org/10.1242/bio.20134614>

- Betancur, R. R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., ... Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, 17(1), 1–40. <http://doi.org/10.1186/s12862-017-0958-3>
- Boc, A., Diallo, A.B., & Makarenkov, V. (2012). T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Research*, 40(W1), 573–579. <http://doi.org/10.1093/nar/gks485>
- Bryant, D. M., & Stow, J. L. (2004). The ins and outs of E-cadherin trafficking. *Trends in Cell Biology*, 14(8), 427–434. <http://doi.org/10.1016/j.tcb.2004.07.007>
- Calvo, F., Ege, N., Grande-Garcia, A., Hooper, S., Jenkins, R. P., Chaudhry, S. I., ... Sahai, E. (2013). Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nature Cell Biology*, 15(6), 637–646. <http://doi.org/10.1038/ncb2756>
- Carreira-Barbosa, F. (2003). Prickle 1 regulates cell movements during gastrulation and neuronal migration in zebrafish. *Development*, 130(17), 4037–4046. <http://doi.org/10.1242/dev.00567>
- Cebola, I., Rodríguez-Seguí, S. a, Cho, C. H.-H., Bessa, J., Rovira, M., Luengo, M., ... Ferrer, J. (2015). TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nature Cell Biology*, 17(5), 615–26. <http://doi.org/10.1038/ncb3160>
- Cheetham, S. W., Gruhn, W. H., van den Ameele, J., Krautz, R., Southall, T. D., Kobayashi, T., ... Brand, A. H. (2018). Targeted DamID reveals differential binding of mammalian pluripotency factors, (September). <http://doi.org/10.1242/dev.170209>
- Chiou, M. J., Chao, T. T., Wu, J. L., Kuo, C. M., & Chen, J. Y. (2006). The physiological role of CTGF/CCN2 in zebrafish notochord development and biological analysis of the proximal promoter region. *Biochemical and Biophysical Research Communications*, 349(2), 750–758. <http://doi.org/10.1016/j.bbrc.2006.08.095>
- Choi, S. C., & Han, J. K. (2002). *Xenopus* Cdc42 regulates convergent extension movements during gastrulation through Wnt/Ca²⁺-signaling pathway. *Developmental Biology*, 244(2), 342–357. <http://doi.org/10.1006/dbio.2002.0602>

- Conway, G., Margoliath, a, Wong-Madden, S., Roberts, R. J., & Gilbert, W. (1997). Jak1 kinase is required for cell migrations and anterior specification in zebrafish embryos. *Proceedings of the National Academy of Sciences of the United States of America*, 94(7), 3082–7. <http://doi.org/10.1073/pnas.94.7.3082>
- Coyle, R. C., Latimer, A., & Jessen, J. R. (2008). Membrane-type 1 matrix metalloproteinase regulates cell migration during zebrafish gastrulation: Evidence for an interaction with non-canonical Wnt signaling. *Experimental Cell Research*, 314(10), 2150–2162. <http://doi.org/10.1016/j.yexcr.2008.03.010>
- Dai, X., She, P., Chi, F., Feng, Y., Liu, H., Jin, D., ... Zhao, B. (2013). Phosphorylation of angiomin by Lats1/2 kinases inhibits F-actin binding, cell migration, and angiogenesis. *The Journal of Biological Chemistry*, 288(47), 34041–51. <http://doi.org/10.1074/jbc.M113.518019>
- Davidson, L. A., Marsden, M., Keller, R., & DeSimone, D. W. (2006). Integrin $\alpha 5 \beta 1$ and Fibronectin Regulate Polarized Cell Protrusions Required for Xenopus Convergence and Extension. *Current Biology*, 16(9), 833–844. <http://doi.org/10.1016/j.cub.2006.03.038>
- Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology*, 3(10). <http://doi.org/10.1371/journal.pbio.0030314>
- Deng, X., & Fang, L. (2018). VGLL4 is a transcriptional cofactor acting as a novel tumor suppressor via interacting with TEADs, 8(6), 932–943.
- Deng, Y., Wu, L. M. N., Bai, S., Zhao, C., Wang, H., Wang, J., ... Lu, Q. R. (2017). A reciprocal regulatory loop between TAZ/YAP and G-protein Gas regulates Schwann cell proliferation and myelination. *Nature Communications*, 8, 1–15. <http://doi.org/10.1038/ncomms15161>
- Dohn, M. R., Mundell, N. A., Sawyer, L. M., Dunlap, J. A., & Jessen, J. R. (2013). Planar cell polarity proteins differentially regulate extracellular matrix organization and assembly during zebrafish gastrulation. *Developmental Biology*, 383(1), 39–51. <http://doi.org/10.1016/j.ydbio.2013.08.027>
- Dupont, S., Morsut, L., Aragona, M., Enzo, E., Giulitti, S., Cordenonsi, M., ... Piccolo, S. (2011). Role of YAP/TAZ in mechanotransduction. *Nature*, 474(7350), 179–83. <http://doi.org/10.1038/nature10137>

- Elbediwy, A., & Thompson, B. J. (2018). Evolution of mechanotransduction via YAP/TAZ in animal epithelia. *Current Opinion in Cell Biology*, 51, 117–123. <http://doi.org/10.1016/j.ceb.2018.02.003>
- Elosegui-Artola, A., Andreu, I., Beedle, A. E. M., Lezamiz, A., Uroz, M., Kosmalka, A. J., ... Roca-Cusachs, P. (2017). Force Triggers YAP Nuclear Entry by Regulating Transport across Nuclear Pores. *Cell*, 171(6), 1397–1410.e14. <http://doi.org/10.1016/j.cell.2017.10.008>
- Elosegui-Artola, A., Oria, R., Chen, Y., Kosmalka, A., Pérez-González, C., Castro, N., ... Roca-Cusachs, P. (2016). Mechanical regulation of a molecular clutch defines force transmission and transduction in response to matrix rigidity. *Nature Cell Biology*, 18(5), 540–548. <http://doi.org/10.1038/ncb3336>
- Engler, A. J., Sen, S., Sweeney, H. L., & Discher, D. E. (2006). Matrix elasticity directs stem cell lineage specification. *Cell*, 126(4), 677–89. <http://doi.org/10.1016/j.cell.2006.06.044>
- Estarás, C., Hsu, H. T., Huang, L., & Jones, K. A. (2017). YAP repression of the WNT3 gene controls hESC differentiation along the cardiac mesoderm lineage. *Genes and Development*, 31(22), 2250–2263. <http://doi.org/10.1101/gad.307512.117>
- Eyckmans, J., Boudou, T., Yu, X., & Chen, C. S. (2011). A hitchhiker's guide to mechanobiology. *Developmental Cell*, 21(1), 35–47. <http://doi.org/10.1016/j.devcel.2011.06.015>
- Feldman, B., Gates, M. A., Egan, E. S., Dougan, S. T., Rennebeck, G., Sirotkin, H. I., ... Talbot, W. S. (1998). Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature*, 395(6698), 181–185. <http://doi.org/10.1038/26013>
- Ferrigno, O., Lallemand, F., Verrecchia, F., L'hoste, S., Camonis, J., Atfi, A., & Mauviel, A. (2002). Yes-associated protein (YAP65) interacts with Smad7 and potentiates its inhibitory activity against TGF- β /Smad signaling. *Oncogene*, 21(32), 4879–4884. <http://doi.org/10.1038/sj.onc.1205623>
- Fukui, H., Terai, K., Nakajima, H., Chiba, A., Fukuhara, S., & Mochizuki, N. (2014). S1P-Yap1 signaling regulates endoderm formation required for cardiac precursor cell migration in zebrafish. *Developmental Cell*, 31(1), 128–136. <http://doi.org/10.1016/j.devcel.2014.08.014>
- Furthauer, M. (2004). Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. *Development*, 131(12), 2853–2864. <http://doi.org/10.1242/dev.01156>

- Furushima, K., Yamamoto, A., Nagano, T., Shibata, M., Miyachi, H., Abe, T., ... Aizawa, S. (2007). Mouse homologues of Shisa antagonistic to Wnt and Fgf signalings. *Developmental Biology*, 306(2), 480–492. <http://doi.org/10.1016/j.ydbio.2007.03.028>
- Furutani-Seiki, M., Sasado, T., Morinaga, C., Suwa, H., Niwa, K., Yoda, H., ... Kondoh, H. (2004). A systematic genome-wide screen for mutations affecting organogenesis in Medaka, *Oryzias latipes*. *Mechanisms of Development*, 121(7–8), 647–658. <http://doi.org/10.1016/j.mod.2004.04.016>
- Galli, G. G., Carrara, M., Yuan, W. C., Valdes-Quezada, C., Gurung, B., Pepe-Mooney, B., ... Camargo, F. D. (2015). YAP Drives Growth by Controlling Transcriptional Pause Release from Dynamic Enhancers. *Molecular Cell*, 60(2), 328–337. <http://doi.org/10.1016/j.molcel.2015.09.001>
- Gorelik, R., & Gautreau, A. (2014). Quantitative and unbiased analysis of directional persistence in cell migration. *Nature Protocols*, 9(8), 1931–1943. <http://doi.org/10.1038/nprot.2014.131>
- Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, 27(7), 1017–1018. <http://doi.org/10.1093/bioinformatics/btr064>
- Gutierrez-Triana, J. A., Mateo, J. L., Ibberson, D., Ryu, S., & Wittbrodt, J. (2016). iDamIDseq and iDEAR: an improved method and computational pipeline to profile chromatin-binding proteins. *Development*, 143(22), 4272–4278. <http://doi.org/10.1242/dev.139261>
- Habas, R., Dawid, I. B., & He, X. (2003). Coactivation of Rac and Rho by Wnt/Frizzled signaling is required for vertebrate gastrulation. *Genes and Development*, 17(2), 295–309. <http://doi.org/10.1101/gad.1022203>
- Harvey, K. F., Pflieger, C. M., & Hariharan, I. K. (2003). The Drosophila Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. *Cell*, 114(4), 457–467. [http://doi.org/10.1016/S0092-8674\(03\)00557-9](http://doi.org/10.1016/S0092-8674(03)00557-9)
- Heisenberg, C.-P., Tada, M., Rauch, G.-J., Saude, L., Concha, M. L., Geisler, R., ... Wilson, S. W. (2000). Silberblick / Wnt11 mediates convergent extension movements during zebra [®] sh gastrulation. *Nature*, 405(May), 76–81. <http://doi.org/10.1038/35011068>

- Hilman, D., & Gat, U. (2011). The evolutionary history of YAP and the Hippo/YAP pathway. *Molecular Biology and Evolution*, 28(8), 2403–2417.
<http://doi.org/10.1093/molbev/msr065>
- Hoch, R. V. (2003). Roles of PDGF in animal development. *Development*, 130(20), 4769–4784.
<http://doi.org/10.1242/dev.00721>
- Huang, J., Wu, S., Barrera, J., Matthews, K., & Pan, D. (2005). The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila homolog of YAP. *Cell*, 122(3), 421–434.
<http://doi.org/10.1016/j.cell.2005.06.007>
- Iioka, H., Ueno, N., & Kinoshita, N. (2004). Essential role of MARCKS in cortical actin dynamics during gastrulation movements. *Journal of Cell Biology*, 164(2), 169–174.
<http://doi.org/10.1083/jcb.200310027>
- Iwamatsu, T. (2004). Stages of normal development in the medaka *Oryzias latipes*. *Mechanisms of Development*, 121(7–8), 605–618.
<http://doi.org/10.1016/j.mod.2004.03.012>
- Jopling, C., & den Hertog, J. (2005). Fyn/Yes and non-canonical Wnt signalling converge on RhoA in vertebrate gastrulation cell movements. *EMBO Rep*, 6(5), 426–431.
<http://doi.org/10.1038/sj.embor.7400386>
- Kanai, F., Marignani, P. A., Sarbassova, D., Yagi, R., Hall, R. A., Donowitz, M., ... Yaffe, M. B. (2000). TAZ: A novel transcriptional co-activator regulated by interactions with 14-3-3 and PDZ domain proteins. *EMBO Journal*, 19(24), 6778–6791.
<http://doi.org/10.1093/emboj/19.24.6778>
- Kane, D., & Adams, R. (2002). Life at the edge: epiboly and involution in the zebrafish. *Results and Problems in Cell Differentiation*, 40, 117–135
- Keller, R. E. (1980). The cellular basis of epiboly: an SEM study of deep-cell rearrangement during gastrulation in *Xenopus laevis*. *Journal of Embryology and Experimental Morphology*, 60(1939), 201–234.
- Keller, R. E., Danilchik, M., Gimlich, R., & Shih, J. (1985). The function and mechanism of convergent extension during gastrulation of *Xenopus laevis*. *Journal of Embryology and Experimental Morphology*, 89, 185–209.

- Keller, P. J., Schmidt, A. D., Wittbrodt, J., & Stelzer, E. H. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science*, 322(5904), 1065-9. doi: 10.1126/science.1162493
- Kikuchi, Y., Verkade, H., Reiter, J. F., Kim, C. H., Chitnis, A. B., Kuroiwa, A., & Stainier, D. Y. R. (2004). Notch Signaling Can Regulate Endoderm Formation in Zebrafish. *Developmental Dynamics*, 229(4), 756–762. <http://doi.org/10.1002/dvdy.10483>
- Kilian, B., Mansukoski, H., Barbosa, F. C., Ulrich, F., Tada, M., & Heisenberg, C. P. (2003). The role of Ppt/Wnt5 in regulating cell shape and movement during zebrafish gastrulation. *Mechanisms of Development*, 120(4), 467–476. [http://doi.org/10.1016/S0925-4773\(03\)00004-2](http://doi.org/10.1016/S0925-4773(03)00004-2)
- Kim, N. G., & Gumbiner, B. M. (2015). Adhesion to fibronectin regulates Hippo signaling via the FAK-Src-PI3K pathway. *Journal of Cell Biology*, 210(3), 503–515. <http://doi.org/10.1083/jcb.201501025>
- Kimelman, D., Smith, N. L., Lai, J. K. H., & Stainier, D. Y. R. (2017). Regulation of posterior body and epidermal morphogenesis in zebrafish by localized Yap1 and Wwtr1. *ELife*, 6, 1–29. <http://doi.org/10.7554/eLife.31065>
- Kimmel, C. B., & Law, R. D. (1985). Cell lineage of zebrafish blastomeres. II. Formation of the yolk syncytial layer. *Developmental Biology*, 108(1), 86–93. [http://doi.org/10.1016/0012-1606\(85\)90011-9](http://doi.org/10.1016/0012-1606(85)90011-9)
- Kimmel, C. B., Warga, R. M., & Schilling, T. F. (1990). Origin and organization of the zebrafish fate map. *Development (Cambridge, England)*, 108(4), 581–594. <http://doi.org/10.1016/j.devcel.2009.04.007>
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics : An Official Public*, 203(3), 253–310. <http://doi.org/10.1002/aja.1002030302>
- Kinoshita, N., Iioka, H., Miyakoshi, A., & Ueno, N. (2003). PKC δ is essential for Dishevelled function in a noncanonical Wnt pathway that regulates *Xenopus* convergent extension movements. *Genes and Development*, 17(13), 1663–1676. <http://doi.org/10.1101/gad.1101303>

- Koos, D. S., & Ho, R. K. (1998). The *nieuwkoid* gene characterizes and mediates a Nieuwkoop-center-like activity in the zebrafish. *Current Biology*, 8(22), 1199–1206. [http://doi.org/10.1016/S0960-9822\(07\)00509-X](http://doi.org/10.1016/S0960-9822(07)00509-X)
- Krieg, M., Arboleda-Estudillo, Y., Puech, P. H., Käfer, J., Graner, F., Müller, D. J., & Heisenberg, C. P. (2008). Tensile forces govern germ-layer organization in zebrafish. *Nature Cell Biology*, 10(4), 429–436. <http://doi.org/10.1038/ncb1705>
- Kuhl, M., Sheldahl, L. C., Park, M., Miller, J. R., & Moon, R. T. (2000). The Wnt/Ca²⁺ pathway A new vertebrate Wnt signaling pathway takes shape. *Trends in Genetics*, 16(7), 279–283. [http://doi.org/10.1016/S0168-9525\(00\)02028-X](http://doi.org/10.1016/S0168-9525(00)02028-X)
- Langdon, Y. G., & Mullins, M. C. (2011). Maternal and Zygotic Control of Zebrafish Dorsoventral Axial Patterning. *Annual Review of Genetics*, 45(1), 357–377. <http://doi.org/10.1146/annurev-genet-110410-132517>
- Latinkic, B. V., Mercurio, S., Bennett, B., Hirst, E. M. a, Xu, Q., Lau, L. F., ... Smith, J. C. (2003). *Xenopus* Cyr61 regulates gastrulation movements and modulates Wnt signalling. *Development* (Cambridge, England), 130(11), 2429–2441. <http://doi.org/10.1242/dev.00449>
- Latimer, A., & Jessen, J. R. (2010). Extracellular matrix assembly and organization during zebrafish gastrulation. *Matrix Biology*, 29(2), 89–96. <http://doi.org/10.1016/j.matbio.2009.10.002>
- Lee, D. H., Park, J. O., Kim, T. S., Kim, S. K., Kim, T. H., Kim, M. C., ... Lim, D. S. (2016). LATS-YAP/TAZ controls lineage specification by regulating TGF β signaling and Hnf4 α expression during liver development. *Nature Communications*, 7(May), 1–14. <http://doi.org/10.1038/ncomms11961>
- Levy, D., Adamovich, Y., Reuven, N., & Shaul, Y. (2008). Yap1 Phosphorylation by c-Abl Is a Critical Step in Selective Activation of Proapoptotic Genes in Response to DNA Damage. *Molecular Cell*, 29(3), 350–361. <http://doi.org/10.1016/j.molcel.2007.12.022>
- Li, J., Chen, X., Ding, X., Cheng, Y., Zhao, B., Lai, Z. C., ... Wang, C. Y. (2013). LATS2 suppresses oncogenic Wnt signaling by disrupting β -Catenin/BCL9 interaction. *Cell Reports*, 5(6), 1650–1663. <http://doi.org/10.1016/j.celrep.2013.11.037>

- Li, P., Silvis, M. R., Honaker, Y., Lien, W. H., Arron, S. T., & Vasioukhin, V. (2016). α E-catenin inhibits a Src–YAP1 oncogenic module that couples tyrosine kinases and the effector of hippo signaling pathway. *Genes and Development*, 30(7), 798–811. <http://doi.org/10.1101/gad.274951.115>
- Li, Z., Wang, Y., Zhang, M., Xu, P., Huang, H., Wu, D., & Meng, A. (2012). The Amotl2 gene inhibits Wnt/ β -catenin signaling and regulates embryonic development in zebrafish. *Journal of Biological Chemistry*, 287(16), 13005–13015. <http://doi.org/10.1074/jbc.M112.347419>
- Li, Z., Zhao, B., Wang, P., Chen, F., Dong, Z., Yang, H., ... Xu, Y. (2010). Structural insights into the YAP and TEAD complex. *Genes & Development*, 24(3), 235–40. <http://doi.org/10.1101/gad.1865810>
- Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., ... Guan, K. (2010). The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation, 1106–1118. <http://doi.org/10.1101/gad.1903310.induces>
- Lin, C., Yao, E., Zhang, K., Jiang, X., Croll, S., Thompson-Peer, K., & Chuang, P. T. (2017). YAP is essential for mechanical force production and epithelial cell proliferation during lung branching morphogenesis. *ELife*, 6, 1–25. <http://doi.org/10.7554/eLife.21130>
- Liu, X., Ma, Y., Zhang, C., Wei, S., Cao, Y., & Wang, Q. (2013). Nodal promotes mir206 expression to control convergence and extension movements during zebrafish gastrulation. *Journal of Genetics and Genomics*, 40(10), 515–521. <http://doi.org/10.1016/j.jgg.2013.07.001>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <http://doi.org/10.1186/s13059-014-0550-8>
- Lynch, C. C., & Matrisian, L. M. (2002). Matrix metalloproteinases in tumor-host cell communication. *Differentiation*, 70(9–10), 561–573. <http://doi.org/10.1046/j.1432-0436.2002.700909.x>
- Malik, A. R., Liszewska, E., & Jaworski, J. (2015). Matricellular proteins of the Cyr61/CTGF/NOV (CCN) family and the nervous system. *Frontiers in Cellular Neuroscience*, 9(June), 1–13. <http://doi.org/10.3389/fncel.2015.00237>

- Mao, Y., & Schwarzbauer, J. E. (2005). Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biology*, 24(6), 389–399. <http://doi.org/10.1016/j.matbio.2005.06.008>
- Marlow, F. L., Marlow, F., Topczewski, J., Sepich, D., & Solnica-krezel, L. (2002). Zebrafish Rho Kinase 2 Acts Downstream of Wnt11 to Mediate Cell Polarity and Effective Convergence and Extension Movements Zebrafish Rho Kinase 2 Acts Downstream of Wnt11 to Mediate Cell Polarity and Effective Convergence and Extension Movements, 12(September 2016), 876–884. [http://doi.org/10.1016/S0960-9822\(02\)00864-3](http://doi.org/10.1016/S0960-9822(02)00864-3)
- Marshall, O. J., & Brand, A. H. (2017). Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. *Nature Communications*, 8(1), 1–9. <http://doi.org/10.1038/s41467-017-02385-4>
- Marshall, O. J., Southall, T. D., Cheetham, S. W., & Brand, A. H. (2016). Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nature Protocols*, 11(9), 1586–1598. <http://doi.org/10.1038/nprot.2016.084>
- McBeath, R., Pirone, D. M., Nelson, C. M., Bhadriraju, K., & Chen, C. S. (2004). Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment. *Developmental Cell*, 6(4), 483–95. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15068789>
- McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis : a unified framework and an evaluation on ChIP data Motif Enrichment Analysis : a unified framework and an evaluation on ChIP data. <http://doi.org/10.1186/1471-2105-11-165>
- Meng, Z., Moroishi, T., & Guan, K. (2016). Mechanisms of Hippo pathway regulation. *Genes and Development*, 30(1), 1–17. <http://doi.org/10.1101/gad.274027.115.1/2>
- Merlot, S. (2003). Leading the way: directional sensing through phosphatidylinositol 3-kinase and other signaling pathways. *Journal of Cell Science*, 116(17), 3471–3478. <http://doi.org/10.1242/jcs.00703>
- Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the panther classification system. *Nature Protocols*, 8(8), 1551–1566. <http://doi.org/10.1038/nprot.2013.092>

- Miesfeld, J. B., Gestri, G., Clark, B. S., Flinn, M. a., Poole, R. J., Bader, J. R., ... Link, B. a. (2015). Yap and Taz regulate retinal pigment epithelial cell fate. *Development*, (July), 3021–3032. <http://doi.org/10.1242/dev.119008>
- Mokalled, M. H., Patra, C., Dickson, A. L., Endo, T., Stainier, D. Y. R., & Poss, K. D. (2016). Injury-induced ctgfa directs glial bridging and spinal cord regeneration in zebrafish, 354(6312), 630–634. <http://doi.org/10.1038/ncomms5234.SUMO1>
- Montero, J.-A. (2005). Shield formation at the onset of zebrafish gastrulation. *Development*, 132(6), 1187–1198. <http://doi.org/10.1242/dev.01667>
- Montero, J. A., Kilian, B., Chan, J., Bayliss, P. E., & Heisenberg, C. P. (2003). Phosphoinositide 3-kinase is required for process outgrowth and cell polarization of gastrulating mesendodermal cells. *Current Biology*, 13(15), 1279–1289
- Moon, R. T., Kohn, A. D., De Ferrari, G. V., & Kaykas, A. (2004). WNT and β -catenin signalling: Diseases and therapies. *Nature Reviews Genetics*, 5(9), 691–701. <http://doi.org/10.1038/nrg1427>
- Moreno-mateos, M. A., Vejnar, C. E., Beaudoin, J., Juan, P., Mis, E. K., Khokha, M. K., ... Haven, N. (2016). HHS Public Access, 12(10), 982–988. <http://doi.org/10.1038/nmeth.3543.CRISPRscan>
- Morin-Kensicki, E. M., Boone, B. N., Howell, M., Stonebraker, J. R., Teed, J., Alb, J. G., ... Milgram, S. L. (2006). Defects in Yolk Sac Vasculogenesis, Chorioallantoic Fusion, and Embryonic Axis Elongation in Mice with Targeted Disruption of Yap65. *Molecular and Cellular Biology*, 26(1), 77–87. <http://doi.org/10.1128/MCB.26.1.77-87.2006>
- Murai, K. K., & Pasquale, E. B. (2005). New exchanges in Eph-dependent growth cone dynamics. *Neuron*, 46(2), 161–163. <http://doi.org/10.1016/j.neuron.2005.04.004>
- Myers, D. C., Sepich, D. S., & Solnica-Krezel, L. (2002). Bmp activity gradient regulates convergent extension during zebrafish gastrulation. *Developmental Biology*, 243(1), 81–98. <http://doi.org/10.1006/dbio.2001.0523>
- Nair, S., & Schilling, T. F. (2008). Chemokine signaling controls endodermal migration during zebrafish gastrulation. *Science*, 322(5898), 89–92. doi: 10.1126/science.1160038

- Nardone, G., Oliver-De La Cruz, J., Vrbsky, J., Martini, C., Pribyl, J., Skládal, P., ... Forte, G. (2017). YAP regulates cell mechanics by controlling focal adhesion assembly. *Nature Communications*, 8(May). <http://doi.org/10.1038/ncomms15321>
- Niehrs, C., Keller, R., Cho, K. W. Y., & DeRobertis, E. M. (1993). The homeobox gene goosecoid controls cell migration in *Xenopus* embryos. *Cell*, 72, 491–503.
- Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., ... Sasaki, H. (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Developmental Cell*, 16(3), 398–410. <http://doi.org/10.1016/j.devcel.2009.02.003>
- Pézeron, G., Mourrain, P., Courty, S., Ghislain, J., Becker, T. S., Rosa, F. M., & David, N. B. (2008). Live Analysis of Endodermal Layer Formation Identifies Random Walk as a Novel Gastrulation Movement. *Current Biology*, 18(4), 276–281. <http://doi.org/10.1016/j.cub.2008.01.028>
- Piccolo, S., Dupont, S., & Cordenonsi, M. (2014). The Biology of YAP/TAZ: Hippo Signaling and Beyond. *Physiological Reviews*, 94(4), 1287–1312. <http://doi.org/10.1152/physrev.00005.2014>
- Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S., & Emslie, K. R. (2012). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Analytical Chemistry*, 84(2), 1003–1011. <http://doi.org/10.1021/ac202578x>
- Porazinski, S., Wang, H., Asaoka, Y., Behrndt, M., Miyamoto, T., Morita, H., ... Furutani-Seiki, M. (2015). YAP is essential for tissue tension to ensure vertebrate 3D body shape. *Nature*, 521(7551), 217–221. <http://doi.org/10.1038/nature14215>
- Rauch, G. J., Hammerschmidt, M., Blader, P., Schauerte, H. E., Strähle, U., Ingham, P. W., McMahon, A. P., & Haftter, P. (1997). Wnt5 is required for tail formation in the zebrafish embryo. *Cold Spring Harbour Symposia on Quantitative Biology*, 62, 227–234
- Rauskolb, C., Sun, S., Sun, G., Pan, Y., & Irvine, K. D. (2014). Cytoskeletal Tension Inhibits Hippo Signaling through an Ajuba-Warts Complex. *Cell*, 158(1), 143–156. <http://doi.org/10.1016/j.cell.2014.05.035>

- Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Mark, H., Borisy, G., ... Horwitz, A. R. (2016). Cell Migration : Integrating Signals from Front to Back Published by : American Association for the Advancement of Science Stable URL : <http://www.jstor.org/stable/3835875> Cell Migration : Integrating Signals from Front to Back, 302(5651), 1704–1709.
- Rohde, L. A., & Heisenberg, C. P. (2007). Zebrafish Gastrulation: Cell Movements, Signals, and Mechanisms. *International Review of Cytology*, 261(07), 159–192. [http://doi.org/10.1016/S0074-7696\(07\)61004-3](http://doi.org/10.1016/S0074-7696(07)61004-3)
- Rosenbluh, J., Nijhawani, D., Cox, A. G., Li, X., Neal, J. T., Schafer, E. J., ... Hahn, W. C. (2012). β -Catenin-driven cancers require a YAP1 transcriptional complex for survival and tumorigenesis. *Cell*, 151(7), 1457–1473. <http://doi.org/10.1016/j.cell.2012.11.026>
- Sampath, K., Rubinstein, A. L., Cheng, A. M. S., Liang, J. O., Fekany, K., Solnica-Krezel, L., ... Wright, C. V. E. (1998). Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling. *Nature*, 395(6698), 185–189. <http://doi.org/10.1038/26020>
- Schepis, A., Sepich, D., & Nelson, W. J. (2012). E-catenin regulates cell-cell adhesion and membrane blebbing during zebrafish epiboly. *Development*, 139(3), 537–546. <http://doi.org/10.1242/dev.073932>
- Schier, A. F. (2003). Nodal Signaling in Vertebrate Development. *Annual Review of Cell and Developmental Biology*, 19(1), 589–621. <http://doi.org/10.1146/annurev.cellbio.19.041603.094522>
- Schötz, E. M., Burdine, R. D., Jülicher, F., Steinberg, M. S., Heisenberg, C. P., & Foty, R. A. (2008). Quantitative differences in tissue surface tension influence zebrafish germ layer positioning. *HFSP Journal*, 2(1), 42–56. <http://doi.org/10.2976/1.2834817>
- Schulte-Merker, S., Eeden, van, Halpern, Kimmel, & Nüsslein-Volhard. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development (Cambridge, England)*, 120(4), 1009–1015. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=7600949%5Cnpapers://01628a62-bdc0-47f5-904e-9689163247b2/Paper/p147

- Sepich, D. S., Calmelet, C., Kiskowski, M., & Solnica-Krezel, L. (2005). Initiation of convergence and extension movements of lateral mesoderm during zebrafish gastrulation. *Developmental Dynamics*, 234(2), 279–292. <http://doi.org/10.1002/dvdy.20507>
- Sepich, D. S., Myers, D. C., Short, R., Topczewski, J., Marlow, F., & Solnica-Krezel, L. (2000). Role of the zebrafish trilobite locus in gastrulation movements of convergence and extension. *Genesis*, 27(4), 159–173. [http://doi.org/10.1002/1526-968X\(200008\)27:4<159::AID-GENE50>3.0.CO;2-T](http://doi.org/10.1002/1526-968X(200008)27:4<159::AID-GENE50>3.0.CO;2-T)
- Sepich, D. S., & Solnica-Krezel, L. (2005). Analysis of cell movements in zebrafish embryos: morphometrics and measuring movement of labeled cell populations in vivo. *Methods in Molecular Biology (Clifton, N.J.)*, 294, 211–33. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15576915>
- Sheldahl, L. C., Slusarski, D. C., Pandur, P., Miller, J. R., Kühl, M., & Moon, R. T. (2003). Dishevelled activates Ca²⁺ flux, PKC, and CamKII in vertebrate embryos. *Journal of Cell Biology*, 161(4), 769–777. <http://doi.org/10.1083/jcb.200211094>
- Shimizu, T., Yabe, T., Muraoka, O., Yonemura, S., Aramaki, S., Hatta, K., ... Hibi, M. (2005). E-cadherin is required for gastrulation cell movements in zebrafish. *Mechanisms of Development*, 122(6), 747–763. <http://doi.org/10.1016/j.mod.2005.03.008>
- Singh, A., Ramesh, S., Cibi, D. M., Yun, L. S., Li, J., Li, L., ... Singh, M. K. (2016). Hippo Signaling Mediators Yap and Taz Are Required in the Epicardium for Coronary Vasculature Development. *Cell Reports*, 15(7), 1384–1393. <http://doi.org/10.1016/j.celrep.2016.04.027>
- Slusarski, D. C., Yang-Snyder, J., Busa, W. B., & Moon, R. T. (1997). Modulation of embryonic intracellular Ca²⁺ signaling by Wnt-5A. *Developmental Biology*, 182(1), 114–20. <http://doi.org/10.1006/dbio.1996.8463>
- Solnica-Krezel, L. (2005). Conserved patterns of cell movements during vertebrate gastrulation. *Current Biology*, 15(6), 213–228. <http://doi.org/10.1016/j.cub.2005.03.016>
- Solnica-Krezel, L., & Sepich, D. S. (2012). Gastrulation: Making and Shaping Germ Layers. *Annu. Rev. Cell Dev. Biol*, 28(June), 687–717. <http://doi.org/10.1146/annurev-cellbio-092910-154043>

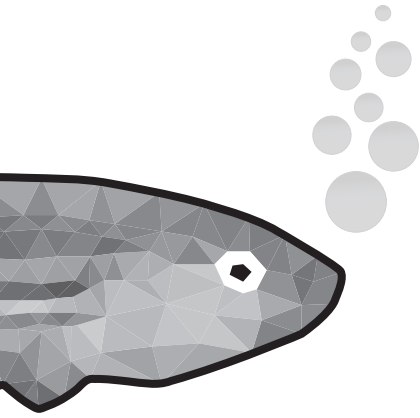
- Song, S., Eckerle, S., Onichtchouk, D., Marrs, J. A., Nitschke, R., & Driever, W. (2013). Pou5f1-Dependent EGF Expression Controls E-Cadherin Endocytosis, Cell Adhesion, and Zebrafish Epiboly Movements. *Developmental Cell*, 24(5), 486–507. <http://doi.org/10.1016/j.devcel.2013.01.016>
- Southall, T. D., Gold, K. S., Egger, B., Davidson, C. M., Caygill, E. E., Marshall, O. J., & Brand, A. H. (2013). Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: Assaying RNA pol II occupancy in neural stem cells. *Developmental Cell*, 26(1), 101–112. <http://doi.org/10.1016/j.devcel.2013.05.020>
- Sprague, J., Clements, D., Conlin, T., Edwards, P., Frazer, K., Schaper, K., ... Westerfield, M. (2003). The Zebrafish Information Network (ZFIN): The zebrafish model organism database. *Nucleic Acids Research*, 31(1), 241–243. <http://doi.org/10.1093/nar/gkg027>
- Stein, C., Bardet, A. F., Roma, G., Bergling, S., Clay, I., Ruchti, A., ... Bauer, A. (2015). YAP1 Exerts Its Transcriptional Control via TEAD-Mediated Activation of Enhancers. *PLoS Genetics*, 11(8), 1–28. <http://doi.org/10.1371/journal.pgen.1005465>
- Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE*, 10(4), 1–11. <http://doi.org/10.1371/journal.pone.0124633>
- Strano, S., Munarriz, E., Rossi, M., Castagnoli, L., Shaul, Y., Sacchi, A., ... Blandino, G. (2001). Physical Interaction with Yes-associated Protein Enhances p73 Transcriptional Activity. *Journal of Biological Chemistry*, 276(18), 15164–15173. <http://doi.org/10.1074/jbc.M010484200>
- Strutt, H., & Strutt, D. (2005). Long-range coordination of planar polarity in *Drosophila*. *BioEssays*, 27(12), 1218–1227. <http://doi.org/10.1002/bies.20318>
- Sudol, M. (1994). Yes-associated protein (YAP65) is a proline-rich phosphoprotein that binds to the SH3 domain of the Yes proto-oncogene product. *Oncogene*, 9(8):2145-2152
- Takeuchi, M., Nakabayashi, J., Sakaguchi, T., Yamamoto, T. S., Takahashi, H., Takeda, H., & Ueno, N. (2003). The prickle-related gene in vertebrates is essential for gastrulation cell movements. *Current Biology*, 13(8), 674-679

- Tamm, C., Bower, N., & Anneren, C. (2011). Regulation of mouse embryonic stem cell self-renewal by a Yes-YAP-TEAD2 signaling pathway downstream of LIF. *Journal of Cell Science*, 124(7), 1136–1144. <http://doi.org/10.1242/jcs.075796>
- Tang, Y., Rowe, R. G., Botvinick, E. L., Kurup, A., Putnam, A. J., Seiki, M., ... Weiss, S. J. (2013). MT1-MMP-dependent control of skeletal stem cell commitment via a β 1-integrin/YAP/TAZ signaling axis. *Developmental Cell*, 25(4), 402–16. <http://doi.org/10.1016/j.devcel.2013.04.011>
- Taniguchi, K., Wu, L. W., Grivennikov, S. I., De Jong, P. R., Lian, I., Yu, F. X., ... Karin, M. (2015). A gp130-Src-YAP module links inflammation to epithelial regeneration. *Nature*, 519(7541), 57–62. <http://doi.org/10.1038/nature14228>
- Tapon, N., Harvey, K. F., Bell, D. W., Wahrer, D. C. R., Schiripo, T. A., Haber, D. A., & Hariharan, I. K. (2002). salvador promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. *Cell*, 110(4), 467–478. [http://doi.org/10.1016/S0092-8674\(02\)00824-3](http://doi.org/10.1016/S0092-8674(02)00824-3)
- Thiery, J. P., Acloque, H., Huang, R. Y. J., & Nieto, M. A. (2009). Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*, 139(5), 871–890. <http://doi.org/10.1016/j.cell.2009.11.007>
- Thisse, C., & Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nature Protocols*, 3(1), 59–69. <http://doi.org/10.1038/nprot.2007.514>
- Topczewski, J., Sepich, D. S., Myers, D. C., Walker, C., Amores, A., Lele, Z., ... Solnica-Krezel, L. (2001). The Zebrafish Glypican Knypek Controls Cell Polarity during Gastrulation Movements of Convergent Extension. *Developmental Cell*, 1(2), 251–264. [http://doi.org/10.1016/S1534-5807\(01\)00005-3](http://doi.org/10.1016/S1534-5807(01)00005-3)
- Tosti, L., Ashmore, J., Tan, B. S. N., Carbone, B., Mistri, T. K., Wilson, V., ... Kaji, K. (2018). Mapping transcription factor occupancy using minimal numbers of cells in vitro and in vivo. *Genome Research*, 28(4), 592–605. <http://doi.org/10.1101/gr.227124.117>
- Trinkaus, J. P., & Lentz, T. L. (1967). Surface specializations of Fundulus cells and their relation to cell movements during gastrulation. *The Journal of Cell Biology*, 32(1), 139–153. <http://doi.org/10.1083/jcb.32.1.139>

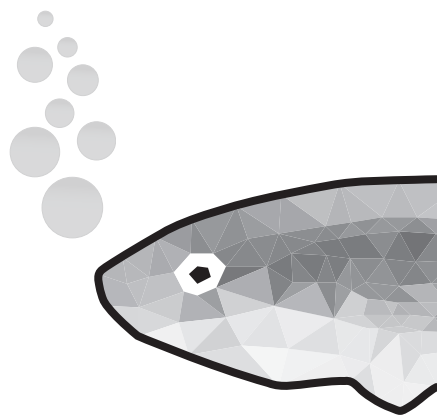
- Ulrich, F., Krieg, M., Schötz, E. M., Link, V., Castanon, I., Schnabel, V., ... Heisenberg, C. P. (2005). Wnt11 functions in gastrulation by controlling cell cohesion through Rab5c and E-Cadherin. *Developmental Cell*, 9(4), 555–564. <http://doi.org/10.1016/j.devcel.2005.08.011>
- van Steensel, B., Delrow, J., & Henikoff, S. (2001). Nature 2001 Chromatin profiling using targeted DNA adenine.pdf, 27(march), 3–7.
- Varelas, X. (2014). The Hippo pathway effectors TAZ and YAP in development, homeostasis and disease. *Development (Cambridge, England)*, 141(8), 1614–26. <http://doi.org/10.1242/dev.102376>
- Varelas, X., Miller, B. W., Sopko, R., Song, S., Gregorieff, A., Fellouse, F. a, ... Attisano, L. (2010). The Hippo pathway regulates Wnt/beta-catenin signaling. *Developmental Cell*, 18(4), 579–91. <http://doi.org/10.1016/j.devcel.2010.03.007>
- Vassilev, A., Kaneko, K. J., Shu, H., Zhao, Y., & Depamphilis, M. L. (2001). TEAD / TEF transcription factors utilize the activation domain of YAP65 , a Src / Yes-associated protein localized in the cytoplasm, 1229–1241. <http://doi.org/10.1101/gad.888601.required>
- Vlahov, N., Scrace, S., Soto, M. S., Grawenda, A. M., Bradley, L., Pankova, D., ... O'Neill, E. (2015). Alternate RASSF1 transcripts control SRC activity, E-cadherin contacts, and YAP-mediated invasion. *Current Biology*, 25(23), 3019–3034. <http://doi.org/10.1016/j.cub.2015.09.072>
- von der Hardt, S., Bakkers, J., Inbal, A., Carvalho, L., Solnica-Krezel, L., Heisenberg, C.-P., & Hammerschmidt, M. (2007). The Bmp gradient of the zebrafish gastrula guides migrating lateral cells by regulating cell-cell adhesion. *Current Biology: CB*, 17(6), 475–87. <http://doi.org/10.1016/j.cub.2007.02.013>
- Wada, K.-I., Itoga, K., Okano, T., Yonemura, S., & Sasaki, H. (2011). Hippo pathway regulation by cell morphology and stress fibers. *Development (Cambridge, England)*, 138(18), 3907–14. <http://doi.org/10.1242/dev.070987>
- Wallingford, J. B., Rowning, B. A., Vogell, K. M., Rothbacher, U., Fraser, S. E., & Harland, R. M. (2000). Dishevelled controls cell polarity during *Xenopus* gastrulation. *Nature*, 405(6782), 81–85. <http://doi.org/10.1038/35011077>

- Warga, R. M., & Kimmel, C. B. (1990). Cell movements during epiboly and gastrulation in zebrafish. *Development*, 108(4), 569–580. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2387236>
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2387236
- Warga, R. M., & Nüsslein-Volhard, C. (1998). spadetail-Dependent cell compaction of the dorsal zebrafish blastula. *Developmental Biology*, 203(1), 116–121. <http://doi.org/10.1006/dbio.1998.9022>
- Westfall, T. A., Hjertos, B., & Slusarski, D. C. (2003). Requirement for intracellular calcium modulation in zebrafish dorsal-ventral patterning. *Developmental Biology*, 259(2), 380–391. [http://doi.org/10.1016/S0012-1606\(03\)00209-4](http://doi.org/10.1016/S0012-1606(03)00209-4)
- Whitman, M. (2001). Nodal Signaling in Early Vertebrate Embryos: Themes and Variations. *Developmental Cell*, 1(5), 605–617. [http://doi.org/10.1016/S1534-5807\(01\)00076-4](http://doi.org/10.1016/S1534-5807(01)00076-4)
- Williams, B. B., Cantrell, V. A., Mundell, N. A., Bennett, A. C., Quick, R. E., & Jessen, J. R. (2012). VANGL2 regulates membrane trafficking of MMP14 to control cell polarity and migration. *Journal of Cell Science*, 125(9), 2141–2147. <http://doi.org/10.1242/jcs.097964>
- Wu, S., Huang, J., Dong, J., & Pan, D. (2003). hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. *Cell*, 114(4), 445–456. [http://doi.org/10.1016/S0092-8674\(03\)00549-X](http://doi.org/10.1016/S0092-8674(03)00549-X)
- Wu, S. Y., Ferkowicz, M., & McClay, D. R. (2007). Ingression of primary mesenchyme cells of the sea urchin embryo: A precisely timed epithelial mesenchymal transition. *Birth Defects Research Part C - Embryo Today: Reviews*, 81(4), 241–252. <http://doi.org/10.1002/bdrc.20113>
- Xin, M., Kim, Y., Sutherland, L. B., Murakami, M., Qi, X., McAnally, J., ... Olson, E. N. (2013). Hippo pathway effector Yap promotes cardiac regeneration. *Proceedings of the National Academy of Sciences*, 110(34), 13839–13844. <http://doi.org/10.1073/pnas.1313192110>
- Yagi, R., Chen, L. F., Shigesada, K., Murakami, Y., & Ito, Y. (1999). A WW domain-containing Yes-associated protein (YAP) is a novel transcriptional co-activator. *EMBO Journal*, 18(9), 2551–2562. <http://doi.org/10.1093/emboj/18.9.2551>

- Yamashita, S., Miyagi, C., Carmany-Rampey, A., Shimizu, T., Fujii, R., Schier, A. F., & Hirano, T. (2002). Stat3 controls cell movements during zebrafish gastrulation. *Developmental Cell*, 2(3), 363–375. [http://doi.org/10.1016/S1534-5807\(02\)00126-0](http://doi.org/10.1016/S1534-5807(02)00126-0)
- Yeo, S. Y., Little, M. H., Yamada, T., Miyashita, T., Halloran, M. C., Kuwada, J. Y., ... Okamoto, H. (2001). Overexpression of a slit homologue impairs convergent extension of the mesoderm and causes cyclopia in embryonic zebrafish. *Developmental Biology*, 230(1), 1–17. <http://doi.org/10.1006/dbio.2000.0105>
- Yu, C., Ji, S. Y., Dang, Y. J., Sha, Q. Q., Yuan, Y. F., Zhou, J. J., ... Fan, H. Y. (2016). Oocyte-expressed yes-associated protein is a key activator of the early zygotic genome in mouse. *Cell Research*, 26(3), 275–287. <http://doi.org/10.1038/cr.2016.20>
- Zaidi, S. K., Sullivan, A. J., Medina, R., Ito, Y., Van Wijnen, A. J., Stein, J. L., ... Stein, G. S. (2004). Tyrosine phosphorylation controls Runx2-mediated subnuclear targeting of YAP to repress transcription. *EMBO Journal*, 23(4), 790–799. <http://doi.org/10.1038/sj.emboj.7600073>
- Zanconato, F., Cordenonsi, M., & Piccolo, S. (2016). YAP/TAZ at the Roots of Cancer. *Cancer Cell*, 29(6), 783–803. <http://doi.org/10.1016/j.ccell.2016.05.005>
- Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., ... Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nature Cell Biology*, 17(9), 1218–27. <http://doi.org/10.1038/ncb3216>
- Zhang, J., Smolen, G. A., & Haber, D. A. (2008). Negative regulation of YAP by LATS1 underscores evolutionary conservation of the Drosophila Hippo pathway. *Cancer Research*, 68(8), 2789–2794. <http://doi.org/10.1158/0008-5472.CAN-07-6205>
- Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., ... Guan, K.-L. (2008). TEAD mediates YAP-dependent gene induction and growth control. *Genes & Development*, 22(14), 1962–71. <http://doi.org/10.1101/gad.1664408>
- Zheng, Y., Wang, W., Liu, B., Deng, H., Uster, E., & Pan, D. (2015). Identification of Happyhour/MAP4K as Alternative Hpo/Mst-like Kinases in the Hippo Kinase Cascade. *Developmental Cell*, 1–14. <http://doi.org/10.1016/j.devcel.2015.08.014>



Anexo



Comparative epigenomics in distantly related teleost species identifies conserved *cis*-regulatory nodes active during the vertebrate phylotypic period

Juan J. Tena,^{1,4} Cristina González-Aguilera,^{1,4} Ana Fernández-Miñán,¹ Javier Vázquez-Marín,¹ Helena Parra-Acero,¹ Joe W. Cross,² Peter W.J. Rigby,² Jaime J. Carvajal,^{1,2} Joachim Wittbrodt,³ José L. Gómez-Skarmeta,^{1,5} and Juan R. Martínez-Morales^{1,5}

¹Centro Andaluz de Biología del Desarrollo (CSIC/UPO/JA), 41013 Sevilla, Spain; ²Division of Cancer Biology, The Institute of Cancer Research, London SW3 6JB, United Kingdom; ³Centre for Organismal Studies, COS, University of Heidelberg, 69120 Heidelberg, Germany

The complex relationship between ontogeny and phylogeny has been the subject of attention and controversy since von Baer's formulations in the 19th century. The classic concept that embryogenesis progresses from clade general features to species-specific characters has often been revisited. It has become accepted that embryos from a clade show maximum morphological similarity at the so-called phylotypic period (i.e., during mid-embryogenesis). According to the hourglass model, body plan conservation would depend on constrained molecular mechanisms operating at this period. More recently, comparative transcriptomic analyses have provided conclusive evidence that such molecular constraints exist. Examining *cis*-regulatory architecture during the phylotypic period is essential to understand the evolutionary source of body plan stability. Here we compare transcriptomes and key epigenetic marks (H3K4me3 and H3K27ac) from medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*), two distantly related teleosts separated by an evolutionary distance of 115–200 Myr. We show that comparison of transcriptome profiles correlates with anatomical similarities and heterochronies observed at the phylotypic stage. Through comparative epigenomics, we uncover a pool of conserved regulatory regions (≈ 700), which are active during the vertebrate phylotypic period in both species. Moreover, we show that their neighboring genes encode mainly transcription factors with fundamental roles in tissue specification. We postulate that these regulatory regions, active in both teleost genomes, represent key constrained nodes of the gene networks that sustain the vertebrate body plan.

[Supplemental material is available for this article.]

Behind the broad anatomical diversity observed in vertebrate species rests a common body plan that is established early during embryogenesis and is shared by the entire clade. Central to our modern view of the ontogeny/phylogeny relationship is the concept that basic animal blueprints stand on the evolutionary conservation of key gene regulatory circuits that define tissue and organ identity during embryogenesis (Davidson and Erwin 2006; Carroll 2008). This notion can be traced back to von Baer's formulations in the 19th century, proposing that embryo development progresses from the more general features of a clade to the specific characters of the species. Or, in other words, that within a particular group the early embryonic forms are more similar than the adults (Gould 1977). During the past few decades, it has become accepted that the window of development at which embryos of a clade show maximum morphological similarity is the phylotypic period (Slack et al. 1993), which does not correspond to the earliest stages of development but rather to mid-embryogenesis once the main body axis has been formed (i.e., pharyngula in vertebrates). However, whether this morphological invariance is also reflected by the conservation of molecular modules has been

the subject of debate. According to the egg-timer/hourglass model, conservation of the body plan would depend on constrained molecular mechanisms operating at the phylotypic phase. Among the potential causative mechanisms postulated are the molecular logic imposed by Hox gene colinearity (Duboule 1994) and the low modularity, and therefore high interdependence, of developmental networks during the phylotypic period (Raff 1996; Galis and Metz 2001). Molecular studies in vertebrates based on the ontogenetic analysis of expression for essential genes, as well as protein–protein interactions and signaling pathways, have failed to identify a clear constrained signature during the phylotypic period, thus supporting a funnel-like model (Roux and Robinson-Rechavi 2008; Comte et al. 2010). However, systematic comparative transcriptomic analyses in vertebrates, *Drosophila*, *Caenorhabditis*, and even in plants have recently provided conclusive evidence for the existence of molecular constraints during mid-embryogenesis. These studies have reported both the convergence of interspecific gene expression and the prevalence of ancient genes at the phylotypic phase (Domazet-Loso and Tautz 2010; Kalinka et al. 2010; Irie and Kuratani 2011; Levin et al. 2012; Quint et al. 2012).

⁴These authors contributed equally to this work.

⁵Corresponding authors

E-mail jlgomska@upo.es

E-mail jrmarmor@upo.es

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.163915.113>.

© 2014 Tena et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Examining the *cis*-regulatory logic is a fundamental step toward understanding the evolutionary sources of the observed developmental constraints imposed on animal body plans. Comparative chromatin immunoprecipitation-sequencing (ChIP-seq) and epigenomics studies have recently opened the possibility of uncovering conserved *cis*-regulatory modules during development (Schmidt et al. 2010; Woo and Li 2012; Xiao et al. 2012; Cotney et al. 2013). In this sense, recent work in zebrafish indicates that enhancers that become activated at late gastrula and remain active during mid-embryogenesis are evolutionarily more conserved than those activated earlier or later during development (Bogdanovic et al. 2012). The direct comparative analysis of functionally conserved enhancers in related species will now shed some light on the constrained architecture of the regulatory networks operating at the phylotypic phase (Nelson and Wardle 2013). To address this issue, we have compared transcriptomes and epigenetic marks from medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*), two distantly related teleosts separated by an evolutionary distance of 115–200 Myr (Furutani-Seiki and Wittbrodt 2004). To complement transcriptomic and epigenomic data sets previously reported in zebrafish (Aday et al. 2011; Bogdanovic et al. 2012; Collins et al. 2012; Pauli et al. 2012; Choudhuri et al. 2013), we have generated RNA-seq and genomic tracks for key histone modifications (H3K4me3 and H3K27ac) from stage 24 (44 hpf) medaka embryos. This embryonic stage in medaka corresponds anatomically to 24-hpf embryos in zebrafish (early pharyngula), that is within the phylotypic period (Kimmel et al. 1995; Iwamatsu 2004). Our comparative analysis of fish transcriptomes shows that expression levels of tissue-specific genes correlate with anatomical similarities and heterochronies between medaka and zebrafish. Furthermore, comparative epigenomic analysis of putative active regulatory regions (PARRs) reveals that only 36% of them (4672 out of 12,938) are conserved at the sequence level between the analyzed teleosts. Among these conserved regions, only 14% (680 out of 4672), here termed shared putative active regulatory regions (SPARRs), are simultaneously active in both species during the phylotypic period. Interestingly, genes associated with this small set of co-acetylated regions show a broader and more complex regulatory landscape. In fact, this collection of genes is highly enriched in transcription factors and signaling molecules with key roles in the control of regulatory circuits involved in the specification of tissues and organs. We propose that SPARRs are evolutionarily constrained nodes that highlight core gene networks involved in the definition of the vertebrate body plan.

Results

Anatomical similarities and heterochronies between zebrafish and medaka phylotypic embryos

The ontogenetic analysis of the cumulative evolutionary age of the zebrafish transcriptome (i.e., age index) has revealed that the most ancient set of transcripts corresponds to the late segmentation to early pharyngula stages. The onset of heart beating and blood circulation at 24 hpf are two prominent morphological features that characterize this period of maximum evolutionary constraint (Domazet-Loso and Tautz 2010). Conveniently for our comparative study, RNA-seq and genomic tracks for H3K4me3 and H3K27ac were previously obtained for 24-hpf zebrafish embryos (Bogdanovic et al. 2012; Collins et al. 2012; Choudhuri et al. 2013). To determine which developmental stage in medaka shows the highest similarity to this zebrafish stage, we examined anatomical

landmarks used as a reference for staging in both species (Kimmel et al. 1995; Iwamatsu 2004). These include, among others, the onset of heart beating and blood circulation, the formation of the optic cup and lens vesicles, or the formation of fin and hepatic buds (Supplemental Table S1). According to most of the features analyzed, medaka embryos show maximum anatomical similarity to 24-hpf zebrafish embryos at ~44–48 h of development (stage 24) (Supplemental Fig. S1). Despite the relatively large evolutionary distance separating both teleost lineages (115–200 Myr), zebrafish and medaka embryos show a very similar body plan within this developmental window (Fig. 1). Therefore, medaka stage 24 and zebrafish 24 hpf were selected as equivalent reference stages in our comparative study.

The relative developmental timing of ontogenetic events is largely conserved between zebrafish and medaka during mid-embryogenesis. This is the case for the onset of heart beating, the development of the optic cup and lens vesicle, and the general morphology of the brain (Fig. 1A–C). In addition to the observed similarities, a few heterochronies (i.e., outliers from the main developmental sequence) were also evident (Fig. 1D; Supplemental Table S1). While somitogenesis has only progressed halfway through in medaka at this stage, it is already completed in zebrafish. Furthermore, in contrast to immobile medaka embryos, zebrafish show spontaneous contractions of the trunk and the tail at 24 hpf (Supplemental Fig. S1; Supplemental Movie 1; Kimmel et al. 1995). This is in agreement with previous observations showing that somitogenesis onset and completion, as well as somite number, vary considerably among vertebrate species (Richardson et al. 1998). A second prominent heterochrony was also noticeable for the formation of the fin buds, which happens much earlier in zebrafish (22 hpf) than in medaka embryos (stage 27) (Fig. 1D). Similarly to somitogenesis, fin bud formation has been described as a developmental process frequently uncoupled from the general zootype in vertebrate embryos (Bininda-Emonds et al. 2007; Sakamoto et al. 2009). Interestingly, we could detect only a couple of anatomical traits for which organogenesis progresses earlier in medaka than in zebrafish: the formation of the hepatic and pancreatic buds (Fig. 1D). This observation is consistent with previous descriptions of endoderm derivatives development in both species (Field et al. 2003; Watanabe et al. 2004).

Tissue-specific expression levels resemble anatomical similarities and heterochronies during the phylotypic period

To examine whether the morphological similarities and asynchronies observed correlate with an underlying molecular activity, we performed RNA sequencing (RNA-seq) analyses in medaka at stage 24 (44 hpf) in duplicate and compared RNA levels with the previously published 24-hpf zebrafish transcriptomes (Collins et al. 2012; Choudhuri et al. 2013). The quality of the medaka RNA-seq data was confirmed by the high correlation of the biological replicates (Pearson correlation = 0.99). Although embryo staging is standardized within the zebrafish community (Kimmel et al. 1995), potential differences in the collection and processing of the embryos may be observed. However, the zebrafish 24-hpf data sets used in this study showed a high Pearson correlation coefficient (0.96), despite having been generated in two independent laboratories (Collins et al. 2012; Choudhuri et al. 2013). For interspecies comparisons, we analyzed the expression levels of a set of 9178 orthologous genes, excluding those with reduced RNA expression (counts per million reads [CPM] < 1) (Supplemental

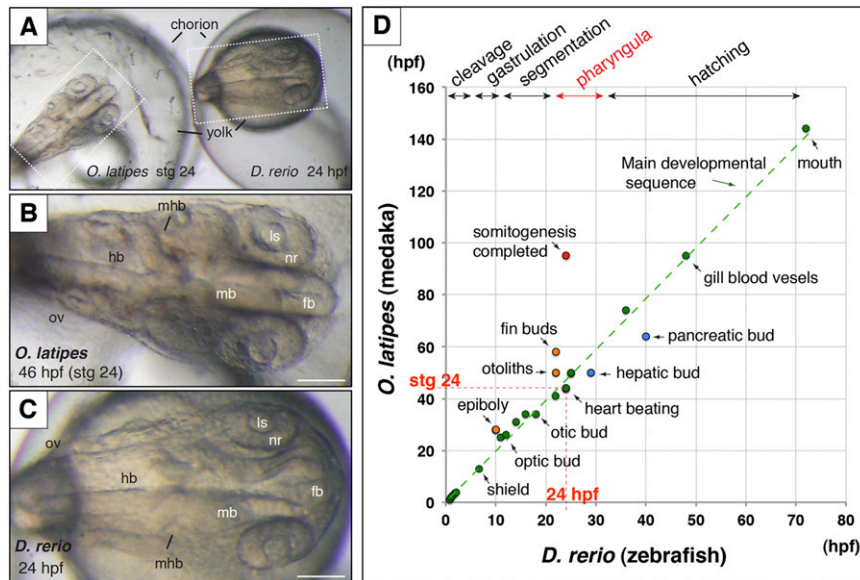


Figure 1. Comparative anatomy of stage 24 (44 hpf) medaka (*O. latipes*) and 24-hpf zebrafish (*D. rerio*) embryos during the phylotypic window (A). Note that both embryo size and general body plan are comparable between medaka (B) and zebrafish (C) at selected stages. The graph shows the onset of key anatomical landmarks plotted in hpf for zebrafish (x-axis) and medaka (y-axis) throughout development (D). The main developmental sequence is indicated as a green dotted line. Red and blue dots represent heterochronic structures between both species. See also Supplemental Table S1. Bar = 100 μ m. (fb) Forebrain, (hb) hindbrain, (ls) lens, (mb) midbrain, (mhb) midbrain-hindbrain boundary, (nr) neural retina, (ov) otic vesicle.

Table S2; see Methods). We found a relatively high correlation between the overall transcriptomes of the two species (Pearson correlation = 0.71) (Fig. 2A). This is in agreement with a previous study comparing vertebrate transcriptomes that shows highest correlation coefficients during the pharyngula window (Irie and Kuratani 2011). To compare gene profiles for different structures, a selection of tissue-specific genes was made based on the ZFIN expression database (Sprague et al. 2006). This list was filtered further through the 9178 orthologous list (Supplemental Table S2). First, we compared the expression levels of genes expressed in the eye, an organ for which no anatomical differences were observed between both species; and in the muscles, for which the differences were evident (Fig. 1). Consistent with the morphological data, we observed that 30% of the genes expressed in the muscles were up-regulated (i.e., more than fourfold) in zebrafish (Fig. 2B). On the contrary, 88.7% of the genes expressed in the eye did not show differential expression between the two organisms (Fig. 2B). We next extended the analysis to other tissues and quantified the significance of the observed changes in expression levels. As is shown in Figure 2C, in addition to the muscles, significant differences in RNA levels were also observed in nervous system-specific genes, which are higher in zebrafish than in medaka. This suggests a premature development of the nervous system in zebrafish that may be consistent with the formation of the neuromuscular junctions required for the active twitching of the tail musculature. With the exception of small but significant differences observed for genes expressed in the epidermis, no additional differences were observed for the rest of the tissue-specific genes examined (Fig. 2C; Supplemental Fig. S2).

To examine whether the divergent expression of tissue-specific genes was due to a delay or an advance in the timing of

ontogenetic events, we included two other reference vertebrates, mouse (*Mus musculus*) and *Xenopus* (*X. tropicalis*) in our comparative transcriptomics analyses. Based on previous studies, we selected embryos within the pharyngula period for these two organisms (Irie and Kuratani 2011). For *Xenopus*, previously published data from stage 24–26 embryos were included in our study (Tan et al. 2013). For mouse, we performed a complete RNA sequencing analysis in duplicate using 10.5-d embryos (Pearson correlation between replicates = 0.75). As expected, pairwise correlation between these four vertebrates revealed that the general expression levels of orthologous genes are more similar in evolutionarily related species (Supplemental Fig. S2A). However, when we analyzed gene expression in specific tissues, we found more similarity when either zebrafish vs. *Xenopus* or medaka vs. mouse were compared (Supplemental Fig. S2C). In particular, transcriptional profiles indicated that specific tissues, such as the muscles and the nervous system, develop comparatively faster in zebrafish and *Xenopus* than in medaka and mouse. A possible explanation for this observation may be derived from

species-specific ecological adaptations during embryogenesis. Zebrafish and *Xenopus* produce large clutches of eggs (100–300 and 1000–3000, respectively) and, most importantly, hatch as free-swimming larvae after a few days of development. In contrast, embryos are produced in smaller numbers (10–30 and 10–15, respectively) and develop at a slower pace in medaka and in the mouse (Supplemental Table S3). This suggests that, although anatomical similarities are maximal at the phylotypic stage, the developmental timing of individual tissues can be conditioned by adaptive requirements and ecological strategies.

To further analyze comparatively the transcriptome of medaka and zebrafish in an independent manner, we computed the number of differentially expressed genes using the edgeR package (Robinson et al. 2010). Selecting a false discovery rate threshold (FDR) < 5% and a fold change greater than fourfold (\log_2 FC > 2), we identified 1085 genes (15.2% of the orthologs list) with higher expression in zebrafish and 600 genes (8.4% of the orthologs) up-regulated in medaka (Supplemental Table S2). Interestingly, the functional categories (Biological process) obtained by DAVID gene ontology (GO) analysis (Huang da et al. 2009) of differentially expressed genes confirmed the up-regulation in zebrafish for muscle tissue development ($P = 5.18 \times 10^{-4}$) and neurological system process- ($P = 1.17 \times 10^{-3}$) related genes. Besides, we found differences in other biological processes not identified through direct morphological observation such as signaling cascade, cardiac muscle tissue development, and protein localization (Fig. 2D; Supplemental Table S2). In the case of medaka up-regulated transcripts, we found that genes related to the cofactor metabolic process ($P = 7.51 \times 10^{-3}$) and oxidation reduction ($P = 4.15 \times 10^{-7}$) were overrepresented with respect to zebrafish. In order to confirm our GO analyses, we decided to use PANTHER, a second bioinformatics tool that has been recently released (Mi et al. 2013).

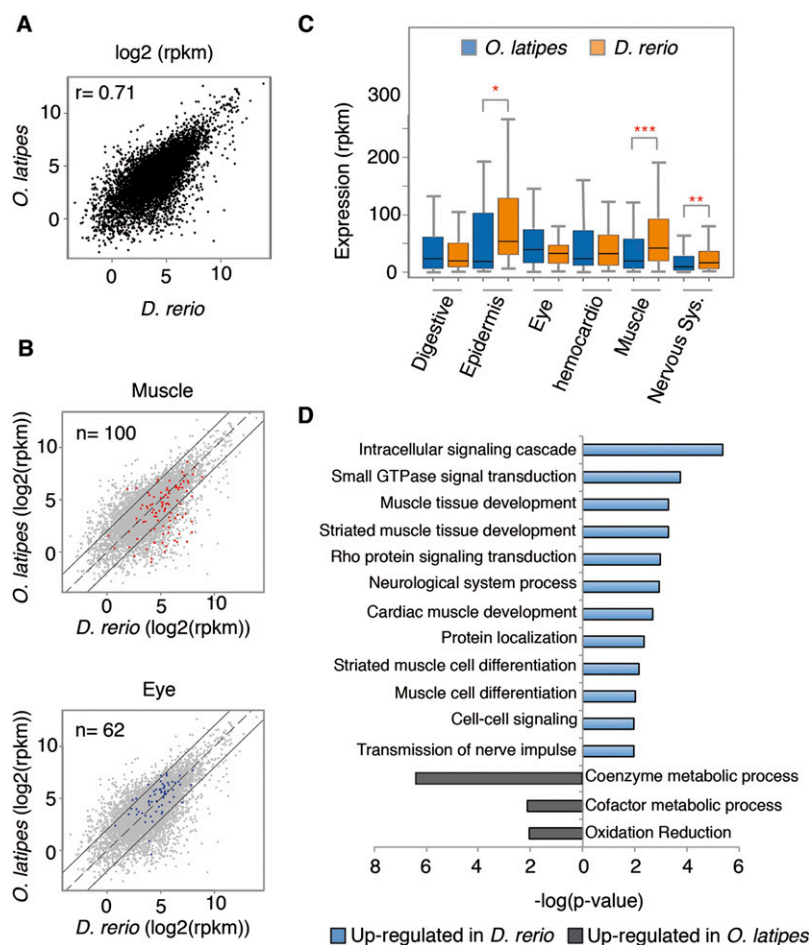


Figure 2. Comparison of expression levels between zebrafish and medaka at the phylotypic stage. (A) Correlation plot of zebrafish and medaka expression levels of the 7118 orthologs with expression higher than 1 count per million reads (CPM). The Pearson correlation coefficient (r) is indicated at the upper left corner. (B) Genes expressed in muscles (red) or the eye (blue), according to ZFIN annotations, are drawn over the total number of orthologous genes (gray). Each point represents the expression level of a given gene (\log_2 RPKM) in both species. Continuous black lines mark a fourfold change in expression. (C) Comparison of the expression levels in different tissues in the two species mentioned above. Bottom and top of boxes indicate 25th and 75th percentiles, respectively, and lines in the boxes indicate medians. Whiskers indicate the lowest and the highest data points within $1.5\times$ interquartile range from the box. (*) $P < 0.05$, (**) $P < 0.01$, (***) $P < 0.001$ (Wilcoxon rank sum test). See also Supplemental Figure S2. (D) Gene ontology categories most significantly overrepresented in differentially expressed genes sorted by P -value. (Blue) Categories up-regulated in zebrafish; (gray) categories up-regulated in medaka.

This second analysis corroborated our previous conclusions, for it showed a significant enrichment in GO terms linked to synaptic transmission and muscle development (e.g., neurological system process, synaptic transmission, mesoderm development, muscle organ development, and transmission of nerve impulse) in genes up-regulated in zebrafish (Supplemental Table S2). In the case of medaka up-regulated transcripts, we found less significantly enriched GO terms, and they were child terms linked to metabolic processes (e.g., lipid metabolic process, cellular amino acid metabolic process, carbohydrate metabolic process), as we observed previously in our DAVID analysis.

All together, these results indicate that the correlation level observed upon comparative analysis of tissue-specific genes resembles not only the anatomical similarities, but also the developmental heterochronies identified between both species.

Identification of conserved H3K27ac marks during the phylotypic period

During embryogenesis, transcriptional control is achieved through the coordinated activation of *cis*-regulatory elements. In recent years, a number of epigenetic marks have been identified as molecular signatures of the activity-state of these regulatory elements (Ong and Corces 2012; Calo and Wysocka 2013). One of them, acetylation of lysine 27 on histone 3 (H3K27ac) has been shown to be a landmark of active transcriptional regulatory elements and promoters in different species (Wang et al. 2008; Heintzman et al. 2009; Creighton et al. 2010; Rada-Iglesias et al. 2011; Bogdanovic et al. 2012). Although comparative analyses of these marks have been performed in a number of cell types, including stem cells (Goke et al. 2011; Woo and Li 2012; Xiao et al. 2012), no such comparisons have been carried out during embryogenesis in general, and at the phylotypic stage in particular. In order to address this point, first we set out to identify active transcriptional regulatory elements in medaka. To that end, we performed ChIP-seq experiments with specific antibodies against H3K4me3 (histone 3 lysine 4 trimethylation) and H3K27ac (Fig. 3A). The reads obtained from sequencing of immunoprecipitated DNA were aligned to the medaka genome (oryLat2 assembly, Ensembl) (Flicek et al. 2013). Then, we used the H3K4me3 mark to filter out promoters from putative active enhancers, both harboring the H3K27ac mark (Fig. 3A C; Ong and Corces 2012). Of 24,027 H3K27ac peaks obtained, we could identify 12,938 that did not overlap with H3K4me3 domains and therefore represent the subset of putative active regulatory regions (PARRs) at this stage (Fig. 3C). The remaining 11,089 H3K27ac peaks represent those regions occupying active

promoters (Fig. 3C). As a validation of our data sets, we found that regions containing both H3K27ac and H3K4me3 marks are associated with transcriptionally active genes, as confirmed by the analysis of our medaka RNA-seq data (Fig. 3D).

Once we identified the putative *cis*-regulatory elements at the phylotypic stage in medaka, we proceeded to analyze their evolutionary conservation, using zebrafish as a reference, a distantly related teleost species. To that end, published ChIP-seq data (Bogdanovic et al. 2012) were used to identify an equivalent set of 8892 PARRs in zebrafish (Supplemental Table S4). For our analyses, we compared these two data sets from medaka and zebrafish, together with a list of conserved regions between both species, as obtained from the UCSC Genome Browser (Meyer et al. 2013). Based on this information, we defined two kinds of conserved DNA domains: Only-one-species PARRs (OPARRs—peaks conserved but

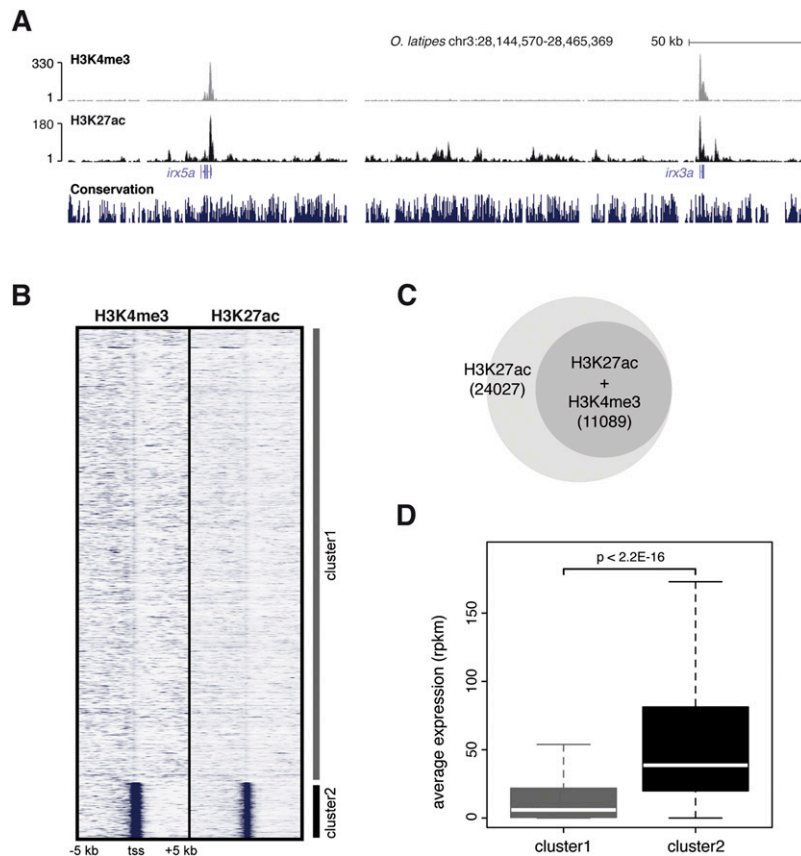


Figure 3. Characterization of epigenetic marks in stage 24 (44 hpf) medaka embryos. (A) UCSC Genome Browser view of H3K4me3 and H3K27ac tracks obtained from medaka ChIP-seq data. As previously described, both epigenomic marks cover the promoter regions of active genes. (B) *K*-means clustering ($k = 2$) of H3K4me3 and H3K27ac signals in ± 5 kb around the TSS of all the genes annotated in the medaka genome. Cluster2 is enriched in both signals around promoters. (C) Venn diagram showing the fraction of H3K27ac regions overlapping with H3K4me3 regions (promoters). (D) Average expression (in RPKM) of genes grouped in clusters in B. Cluster2 genes show an average expression level higher than genes from Cluster1.

putatively active only in one of the two species) and Shared PARRs (SPARRs—peaks conserved and putatively active in both species) (Fig. 4A,B; see peaks validation by qPCR in Methods section). The medaka data set was compared against the zebrafish one, and vice versa. As a result, we obtained 3992 OPARRs and 680 SPARRs in medaka and 2032 OPARRs and 701 SPARRs in zebrafish (Supplemental Table S4). The small discrepancy observed among species in the number of SPARRs is due to both the presence of duplicated regions and the occasional incomplete overlap between SPARRs and conserved regions. To explore the functional significance of these results, we assigned the nearest gene to each PARR to further study their features. Independent of whether gene assignment was examined in medaka or zebrafish, we found that the genomic landscape of SPARR-associated genes had a much wider and higher H3K27ac mean profile than the average of all PARRs-associated genes (Fig. 4C). This might correspond to genes with a high number of *cis*-regulatory regions, many of them located far away from the promoter, which would result in a more complex transcriptional regulation. In fact, when we calculated the number of peaks associated with OPARRs, SPARRs, and all PARRs-related genes, we found that the proportion of genes including several H3K27ac peaks was

significantly higher in the SPARRs subset (Fig. 4D). To minimize potential errors caused by inaccuracies in the assignment of SPARRs to neighbor genes (i.e., due both to local assembly mistakes and to chromosomal rearrangements), we refined our analysis by focusing in on SPARRs associated with the same gene in both species. This refined list of SPARRs, here named as cSPARRs, are associated with genes that showed an even higher number of H3K27ac regulatory regions than the original SPARR-associated genes (Supplemental Fig. S3A,B). Interestingly, a significant fraction of genes associated with SPARRs also include OPARRs in their vicinity, thus indicating that their regulation is more complex (Supplemental Fig. S3C). Taken together, these results indicate that genes with a more complex regulation are also those harboring conserved active enhancers.

To determine whether this conservation is restricted to teleosts or is also maintained in other vertebrates, we compared our data with that of the VISTA Enhancer Browser, a resource including experimentally validated human and mouse noncoding fragments with gene enhancer activity (Visel et al. 2007). In this project, 1857 noncoding human regions selected by means of phylogenetic foot-printing analyses and tissue-specific ChIP-seq assays of epigenomic marks have been tested in mouse transgenic assays. Of these sequences, 982 are able to drive consistent expression patterns and, therefore, are considered as active regulatory regions. Of the 12,938 PARRs

found in medaka, 2157 are conserved with the human genome, and from them 115 overlap with regions analyzed in the VISTA Enhancer Browser collection. A high proportion of these conserved regions (82, 71.3%) were found active in mouse transgenic assays (Supplemental Table S5). When we compared the SPARRs ($n = 680$) from medaka, 253 were conserved in humans. Interestingly, an even higher percentage (88.6%) of the SPARRs were experimentally confirmed as active enhancers in transgenic mice (31 out of the 35 regions found in the VISTA Enhancer Browser database) (Supplemental Fig. S4). Similar results were obtained using the zebrafish data as a reference (Supplemental Table S5). This significantly higher percentage (hypergeometric test, $P = 0.00095$) of positive regulatory regions within the SPARRs suggests that regions putatively active in both teleost species are also active in other vertebrates as well. To test this hypothesis, we crossed the VISTA Enhancer Browser information of elements tested in transgenesis assays with H3K27ac tracks obtained from human ES cells differentiated into distinct cell types representing the basic embryonic cell layers (Xie et al. 2013). Approximately 2/3 (21 out of 31 in medaka and 18 out of 27 in zebrafish) of the regions that showed regulatory activity in mouse transgenesis assays were also acetylated in at least one human differentiated cell type. In con-

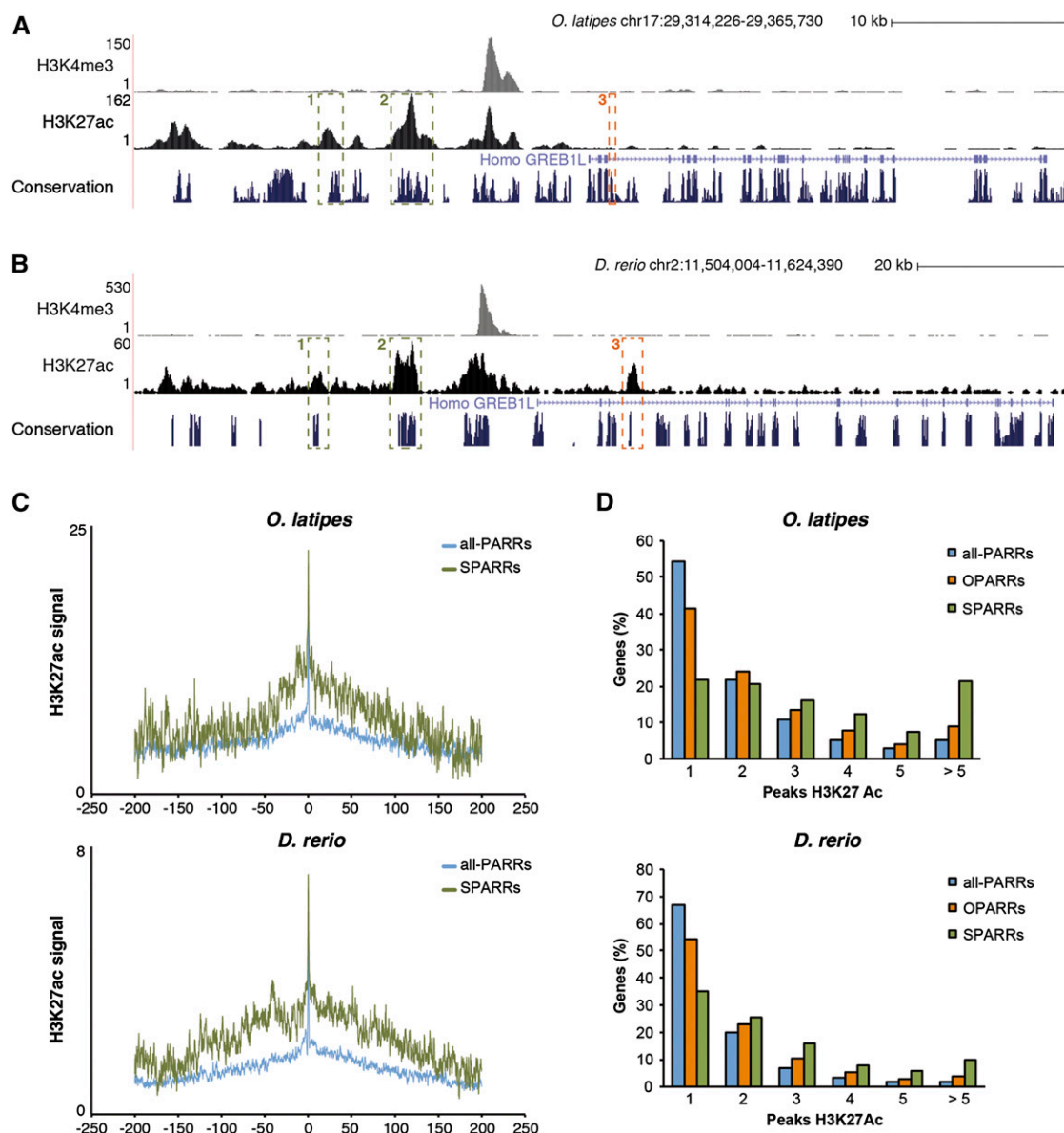


Figure 4. Analysis of the regulatory landscape of phylotypic genes associated with SPARRs and OPARRs. (A,B) Two examples of SPARR (1, 2) and one of OPARR (3) in the medaka genome (A) and their orthologous regions in zebrafish (B) are shown. (C) Average profiles of H3K27ac signal covering a 400-kb landscape for genes associated with all-PARRs (blue line) and SPARRs (green line) in both species: medaka and zebrafish (upper and lower panels, respectively). The average of reads in each bin of 200 bp is represented in log scale on the x-axis. The y-axis shows the position around the gene TSS, in kb. (D) Frequency distribution of orthologous genes (in percent) associated with either all H3K27ac peaks (all-PARRs), OPARRs, or SPARRs, according to the number of H3K27ac regulatory regions included in their vicinity; medaka and zebrafish (upper and lower panels, respectively).

trast, most of the regions (three out of four in medaka and five out of five in zebrafish) that were negative in transgenesis assays were also negative for the acetylation mark in differentiated human ES cells (Supplemental Table S5).

To further validate the functional conservation of SPARRs across vertebrates, we carried out transient transgenesis assays in zebrafish by injecting the corresponding fish regions ($n = 6$) orthologous to the tested mammalian enhancers (VISTA Enhancer Browser). Interestingly, the six regions tested (hs73, hs619, hs625, hs969, hs1315, and hs1327) directed the expression of the reporter GFP in a similar manner (i.e., to the same tissues) as the homologous regions in mice (Supplemental Fig. S5). Moreover, we tested

three of these regions (hs73, hs1315, and hs1327) in transient transgenesis assays in medaka, obtaining very similar results (Supplemental Fig. S5). These results further confirmed that regions active in both teleost species are also functionally conserved in other vertebrates.

Conserved transcriptional control of genes associated with shared regulatory regions

To integrate the information we obtained from the analysis of chromatin epigenetic marks with our gene expression data, we examined the expression levels of genes associated with OPARR

and cSPARRs regions. Expression analysis of medaka genes in the vicinity of H3K27ac regions showed that whereas OPARRs-associated genes display very variable expression levels between species, cSPARR-associated genes were significantly enriched in non-differentially expressed genes ($P = 0.46$ and $P = 0.03$ for OPARR and cSPARRs, respectively; hypergeometric test) (Fig. 5A). Similar results were derived from the analysis of zebrafish genes associated with H3K27ac regions (data not shown). Moreover, we observed that the median expression level of cSPARR-associated genes was significantly higher than the expression average of both genes containing OPARRs and the overall transcriptome (Fig. 5B). These results indicate that the expression control of genes associated with shared regulatory regions is significantly conserved through evolution.

A general DAVID analysis of GO term enrichment in the general list of genes associated with PARRs both in zebrafish and medaka reflected the transcriptionally active state of a broad set of genes related to diverse developmental processes. A number of GO terms involved in tissue patterning (e.g., regionalization, $P = 6.08 \times 10^{-7}$; or pattern specification process, $P = 1.17 \times 10^{-6}$), cellular and epithelial morphogenesis (e.g., tissue morphogenesis, $P = 5.04 \times 10^{-8}$; or cell motion, $P = 3.24 \times 10^{-6}$), or precursor differentiation (e.g., neuron differentiation, $P = 2.13 \times 10^{-5}$) were derived from these analyses (Supplemental Table S6). In contrast, when GO terms were analyzed only for the list of cSPARR-associated

genes, all the significantly enriched terms were related to transcriptional categories such as regulation of transcription: $P = 5.65 \times 10^{-8}$; regulation of RNA metabolism process: $P = 6.04 \times 10^{-8}$; or transcription: $P = 1.57 \times 10^{-4}$ (Fig. 5C; Supplemental Table S6). Moreover, the enrichment analysis of InterPro protein domains related to these cSPARR-associated genes showed the overrepresentation of important transcriptional domains for developmental processes, such as the homeodomain ($P = 7.45 \times 10^{-5}$), zinc finger C2H2 ($P = 8.47 \times 10^{-4}$), SMAD domain ($P = 5.10 \times 10^{-3}$), or winged helix repressor DNA-binding ($P = 9.48 \times 10^{-3}$) (Supplemental Table S6). A detailed analysis of the occurrence of the InterPro domains present in the transcription factors identified within the collection of 145 cSPARR-associated genes is shown in Figure 5D.

Taken together, these results indicate that not only are developmental genes conserved at the vertebrate phylotypic stage (Domazet-Loso and Tautz 2010) but the key regulatory regions responsible for their tight and complex modulation are also conserved. Our data suggest that the shared regulatory elements identified in our study constitute essential nodes of the constrained transcriptional network operating at the phylotypic stage.

Discussion

In this work, we compared zebrafish and medaka pharyngula embryos morphologically and molecularly. We have examined both their transcriptomes and predictive epigenetic marks for conserved active enhancers during the phylotypic window. Whereas in closely related vertebrates, the high overall genome similarity masks the identification of noncoding conserved elements, only a few of them can be identified outside the vertebrate group and even less show enough transphyletic conservation to be tracked beyond the Cambrian horizon (McEwen et al. 2009; Royo et al. 2011; Clarke et al. 2012). The evolutionary distance between zebrafish and medaka (115–200 Myr) is suitable for the identification and analysis of conserved regulatory elements in vertebrates (Furutani-Seiki and Wittbrodt 2004).

Despite their evolutionary distance, zebrafish and medaka share a very similar anatomy, which is particularly noticeable when embryos are compared at the phylotypic stage. Nevertheless, a number of heterochronies are observed during this developmental window (here described in Supplemental Table S1). In fact, the observation of such conspicuous heterochronies between vertebrate phylotypic embryos has been an argument raised against the hourglass model (Richardson et al. 1998). In agreement with the hourglass hypothesis, comparative transcriptomics in vertebrates have revealed that interspecies correlation in gene expression levels is maximal within this phylotypic window (Irie and Kuratani 2011). Our comparative analysis of tissue-

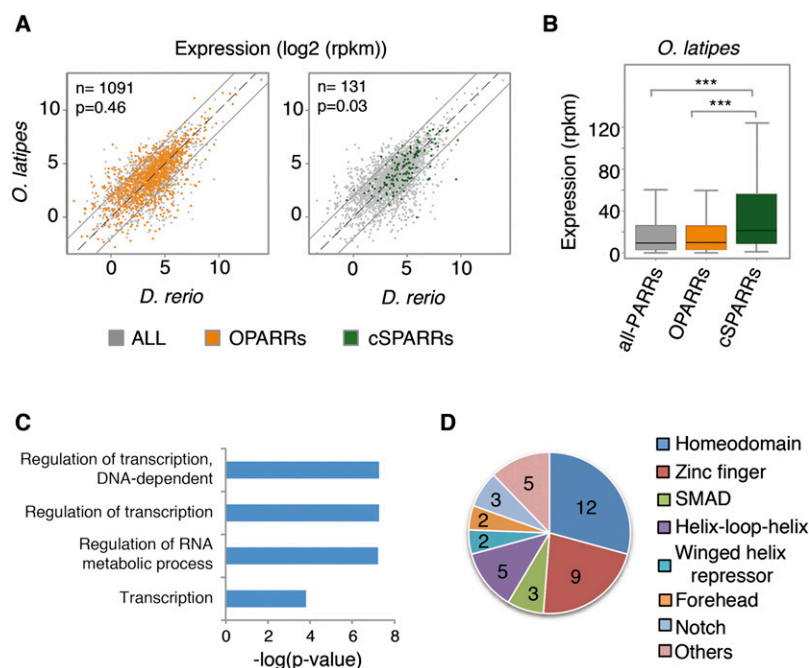


Figure 5. Integration of genome-wide epigenetic and expression data. (A) Comparison of zebrafish and medaka expression levels (log₂ RPKM) for genes associated with OPARR (orange) and cSPARR (green) regions, as identified in medaka. These genes are plotted over the total number of genes associated with H3K27ac regions identified in medaka (gray). The number of genes included in each group is indicated in the upper left corner. Only genes with CPM > 1 are plotted. Note that genes associated with cSPARRs are significantly enriched in nondifferentially expressed transcripts ($P = 0.46$ and $P = 0.03$ for OPARRs and cSPARRs, respectively; hypergeometric test). (B) Box plot indicating the average expression levels (RPKM) of genes associated with cSPARR, OPARR, and all H3K27ac regions identified in medaka. (***) $P < 0.001$ (Wilcoxon rank sum test). (C) Gene ontology analysis of the total number of genes associated with cSPARRs. Significant biological process categories ($P < 0.05$) revealed that genes associated with cSPARRs are involved mainly in transcriptional regulation. (D) Representative InterPro domains present in transcription factors identified within the total collection of genes associated with cSPARRs. Numbers inside the chart indicate the number of genes present in each category.

specific genes shows that there is a high concordance of expression levels in synchronously developing tissues, and thus is in line with previous comparative transcriptomic analyses (Domazet-Loso and Tautz 2010; Irie and Kuratani 2011). In addition, our work shows that this concordance drops when gene expression is compared for heterochronic structures (e.g., muscles and nervous system). This observation fits under the umbrella of the general notion that changes in gene regulatory networks (GRNs) play a prevalent role in the evolution of animal form (Davidson 2006; Carroll 2008).

The objective of this study is not to provide additional evidence showing molecular constraints at the phylotypic period; this has been sufficiently addressed by others. We rather aim to have a first look at the nature of such constraints. *Cis*-regulatory modules (CRMs) have been considered not only the units of input information in GRNs but also the fundamental units of evolutionary change (Davidson 2006). In this report, we have performed a comparative epigenomics study to identify a subset of ~700 putative CRMs (here termed SPARRs) that are both conserved and active in zebrafish and medaka pharyngula embryos. Here we have associated each CRM to the nearest gene. Provided that enhancers for a particular gene could even lie in a neighbor gene intron (Lettice et al. 2003; Smemo et al. 2014), this assumption may lead to potential errors. However, assignment by nearest gene model is the most widely used method, and it has been shown that patterns of enhancer activity correlate strongly with patterns of nearest-gene expression (Ernst et al. 2011; Shen et al. 2012). Our analysis of molecular domains for genes associated with SPARRs reveals that a large proportion of them provide regulatory input to genes encoding transcription factors. This finding suggests that these regulatory regions represent constrained nodes from essential GRNs operating at the phylotypic period in the teleost group. Furthermore, it is likely that the core set of nodes responsible for the evolutionary stability of the vertebrate body plan is, to a large extent, comprised within these regions conserved in teleosts. In agreement with this, a large proportion (88%) of the human SPARRs homologs included in the tested (i.e., in transgenesis assays) collection of CRMs at the VISTA Enhancer Browser behave as tissue-specific active enhancers.

There are a number of reasons to think that the collection of 700 SPARRs identified here represents an underestimate of the actual number of core CRMs responsible for the architecture of the vertebrate phylotype. First, in our analysis we have considered only regulatory regions conserved between the two teleosts, which roughly correspond to a third of the acetylated regions (PARRs) identified in each species. This approach, however, may have excluded a number of elements that still share similar functional logic (i.e., a similar composition of transcription factor binding sites) but whose overall sequence conservation is beyond the detection limits of conventional alignment tools (Fisher et al. 2006; He et al. 2011; Taher et al. 2011). In addition, comparative ChIP-seq studies have also pointed to the existence of pervasive species-specific gene regulation in a number of tissues, including ES cells (The ENCODE Project Consortium 2007; Schmidt et al. 2010; The ENCODE Project Consortium 2012). To what extent this also applies to complex CRMs regulating master developmental genes is currently unclear. Finally, the intrinsic technical limitations imposed by ChIP-seq approaches applied to whole embryos might result in false negatives and hence in an underestimate of the total number of co-acetylated regions in teleost genomes. This may partially explain why a large proportion of the conserved acetylated regions identified in our study appear to be active only in one

species at the phylotypic period (OPARRs). Although we show that, collectively, gene regulatory features associated with SPARRs and OPARRs are significantly different, we cannot rule out the possibility that a fraction of the regions classified here as OPARRs is in fact active below the detection level in one of the species (i.e., due to different regulatory weight). Alternatively, the differential and complex activation timing of these regions in the teleost genomes could also account for the observed prevalence of OPARRs versus SPARRs during the narrow developmental window under study.

Among vertebrate regulatory sequences, evolutionary divergence has been proposed to occur faster in fish genomes. The partitioning of regulatory elements between duplicate gene loci after fish-specific whole-genome duplication (FSGD) has been suggested as a causative mechanism driving their divergence and hence the extensive adaptive radiation observed in teleosts (Taylor et al. 2001; Christoffels et al. 2004; Hoegg et al. 2004; Meyer and Van de Peer 2005). Thus, it has been shown that more than twice as many noncoding elements are conserved between elephant shark and human genomes than between teleost fish and human genomes (Venkatesh et al. 2006). Moreover, comparative genomics studies have shown that conserved noncoding elements have been evolving rapidly in teleost fishes (Wang et al. 2009; Lee et al. 2011). Comparative analyses of epigenetic marks in other vertebrates including tetrapods, cartilaginous fish, and agnates will complement our study and help to define more precisely the ancestral set of CRMs in vertebrates. The analysis of these marks in basal ray-finned fish that diverged from teleosts before the FSGD, such as the spotted gar (*Lepisosteus oculatus*) (Amores et al. 2011), may be also important to determine the degree of regulatory divergence in the teleost group. However, even if our comparative analysis in teleosts overlooks a fraction of the ancestral set of vertebrate CRMs, our approach would be biased toward the identification of “essential nodes,” precisely those enhancers more resilient to evolutionary change due to their central role in the definition of the vertebrate body plan.

It has been shown that although epigenomic conservation does not always correlate with genomic sequence conservation, it can provide an additional layer of information that is necessary to interpret genome regulation (Xiao et al. 2012). Hence, the collection of shared enhancers identified here represents a powerful resource to investigate the architecture of the GRNs operating during the phylotypic window. It has been postulated that the developmental programs controlling different organ primordia may be interdependent in a way that cannot be resolved into individual modules. This lack of modularity may have functioned as an evolutionary constraint to stabilize the vertebrate body plan (Raff 1996). Some of the data presented here are in line with this hypothesis. A large proportion of the SPARR-associated genes encode transcription factors and components of signaling pathways that, in turn, may act as upstream regulators of other conserved nodes of the GRNs. In addition, SPARR-associated genes show a complex regulatory profile, often including multiple CRMs, which suggests that they represent “hub” genes with high connectivity within the GRNs. In fact, an important proportion (between 44%–53%) of the SPARR-associated genes also include in their neighborhood conserved regions that are acetylated only in one of the two species (here termed OPARRs). Whether these putative enhancers act as “shadow enhancers” providing functional robustness to a “primary” enhancer (Hong et al. 2008; Frankel et al. 2010) or, alternatively, bring independent regulatory input needs to be determined. Furthermore, detailed analyses of predicted connectivity focused in the nodes of phylotypic and non-

phylotypic GRNs will be required to formally prove Raff's lack-of-modularity hypothesis.

Methods

Fish stocks and genomes

Medaka (*O. latipes*) and zebrafish (*D. rerio*) wild-type strains were kept as closed stocks, and embryos were staged as previously described (Kimmel et al. 1995; Iwamatsu 2004). The genome assemblies for medaka and zebrafish genomes have been released (Flicek et al. 2013). The zebrafish genome (Zv9) has a size of 1505 Mb, and 26,206 protein-coding genes have been annotated (Collins et al. 2012; Howe et al. 2013). The medaka genome (HdrR-2005) has a size of 700 Mb, and a total of 20,141 coding genes have been predicted (Kasahara et al. 2007).

Embryo collection and RNA samples

Whole medaka and mouse embryos (without extra-embryonic membranes) were collected according to standard procedures. All the animal experiments were carried out in accordance with the guidelines of our Institutional Animal Ethics Committee. For medaka experiments, a total of 60 embryos were suspended in TRIzol reagent (Intron Biotechnology) with chloroform. Two replicates for each sample were used for RNA-seq analyses. Total RNA was isolated from the aqueous phase, purified by isopropanol precipitation, and cleaned using the RNeasy MinElute Cleanup kit (Qiagen). For mouse samples, four day 10.5 embryos were pooled and homogenized, and total RNA was extracted similarly. Two replicates for each sample were also used for mouse RNA-seq analyses. Subsequent processes, including preparation of sequencing libraries, were performed by standard TruSeq™ RNA sample preparation (Illumina) with the following changes: Purifications were carried out using Qiagen clean-up columns, and e-gels were used for size selection. Samples were sequenced using HiSeq 2000 at the EMBL Genomics Core Facility (EMBL-Heidelberg).

Criteria for orthologous genes identification

Basic Local Alignment Search Tool (BLAST) searching ($E\text{-value} < 1 \times 10^{-20}$) was applied to the nonredundant proteome of each organism downloaded from the EMBL Ensembl website (<http://www.ensembl.org/>). Pairs of genes with reciprocal best BLAST hits (RBBHs) were defined to be orthologs.

RNA-seq data processing

Raw RNA sequence data from medaka and mouse, and previously published *Xenopus* (Tan et al. 2013) and zebrafish (Collins et al. 2012; Choudhuri et al. 2013) data, were aligned with the oryLat2 (October 2005), mm9 (July 2007), xenTro2 (August 2005), and danRer7 (July 2010) genome versions, respectively, using TopHat (Trapnell et al. 2009). To minimize errors due to the variability between species annotations, we took into account only the number of mapped reads that overlapped with the Ensembl coding sequences of those genes present in our orthologs list. Expression values were obtained by calculating the sum of all the expression hits from distinct exons annotated to a single locus using RSeQC software (Wang et al. 2012). To filter low-expressed genes, loci with counts per million reads (CPM) < 1 in at least two samples were discarded. For differential expression analyses, raw count data were processed using the edgeR package under default conditions (Robinson and Smyth 2008), and genes with FDR $< 5\%$ and fold change > 4 were considered significant. For analyses of expression

levels in different tissues, data were normalized by scaling read counts to reads per kilobase per million reads (RPKM), followed by quantile normalization to reduce variability between samples. Data were \log_2 transformed and the mean of the replicates was used in further analyses. Genes expressed in specific tissues were obtained from Ensembl filtered by expression in ZFIN (Sprague et al. 2006) anatomical system data: "digestive," "epidermis," "eye," "hemocardio," "muscle," and "nervous," and filtered to obtain genes that only are expressed in one of the tissues (Supplemental Table S2).

Gene ontology analyses

Gene ontology analyses were performed using DAVID (Huang et al. 2009) and PANTHER (Mi et al. 2013). Only the "Biological Process" tree was used in the study. As the medaka genome was not represented in DAVID, only zebrafish Ensembl gene names were used for the analysis. For GO analyses of differentially expressed genes, we used as a reference background the list of orthologous genes with CPM > 1 (Supplemental Table S2). We considered significant GO categories with a $P\text{-value} < 0.05$ and more than seven genes. For GO analyses of ChIP-seq data, the complete list of orthologous genes was used as background. GO categories with a $P\text{-value} < 0.05$ were considered significant. $P\text{-values}$ were corrected by multiple testing.

Medaka ChIP-seq

Chromatin immunoprecipitation (ChIP) was performed following a protocol reported for zebrafish (Bogdanovic et al. 2013) with minor modifications. Per ChIP, we used 600 dechorionated embryos at stage 24. Samples were sonicated using the Diagenode Bioruptor device with the following cycling conditions: 12 min high–30 sec on, 30 sec off; 12 min on ice; 12 min high–30 sec on, 30 sec off. The size of sonicated DNA was in the range of 100–500 bp. The anti-H3K4me3 (pAB-033-050) antibody was obtained from Diagenode. The anti-H3K27ac (ab4729) antibody was purchased from Abcam. Immunoprecipitated DNA was purified with QIAquick columns (Qiagen). DNA ends were repaired and the adaptors ligated. The size-selected (300 bp) library was then amplified in a PCR reaction and sequenced using the Genome Analyzer (Illumina). The sequenced reads were mapped to the reference medaka genome (oryLat2 assembly) with Bowtie software (Langmead et al. 2009). Peak callings were performed with MACS (Zhang et al. 2008) using default parameters. Peaks were independently validated by qPCR, using specific primers for 12 regions, four of them acetylated in medaka but not in zebrafish, four acetylated in zebrafish but not in medaka, and four acetylated in both species (Supplemental Fig. S6). To further test the reproducibility of the ChIP-seq experiment for H3K27ac marks, a second biological replicate was analyzed. Reads from both replicates (grouped in windows of 1 kb over the genome) show a Pearson correlation coefficient of 0.97 (Supplemental Fig. S7).

Comparison between zebrafish and medaka ChIP-seq data

In order to compare acetylation peaks obtained from ChIP-seq analyses in both species, chained and netted alignments (axt format) between danRer7 and oryLat2 assemblies were downloaded from the UCSC Genome Browser downloads web page (<http://hgdownload.cse.ucsc.edu/downloads.html>). Minus-strand coordinates were transformed to plus-strand coordinates. To obtain epigenomic marks corresponding to enhancers, only those H3K27-acetylated peaks that do not overlap with an H3K4-trimethylated peak were considered for each species. This subtraction was per-

formed using BEDTools software (Quinlan and Hall 2010). Putative active regulatory regions identified in one species were crossed with the axt alignment file to map the orthologous positions in the second species. The resulting list was then compared with the mapped PARRs in the second species. A fraction of the PARRs in each species overlapped with conserved regions from the UCSC Genome Browser. For zebrafish PARRs, this overlapping was determined as 86.5% of the total length of the conserved region, on average. For medaka, the mean overlap was 80.5% of the conserved region.

Integration of ChIP-seq and RNA-seq data

RNA-seq profiles were integrated with ChIP-seq data by assigning each acetylated region to its nearest gene using BEDTools. The expression levels of genes associated with PARRs were obtained from our medaka RNA-seq or from the reported zebrafish data sets (Collins et al. 2012; Choudhuri et al. 2013)

Data access

The ChIP-seq and RNA-seq data included in this work have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under the following accession numbers: GSE46351 (medaka stage 24 ChIP-seq tracks), GSE46484 (medaka stage 24 RNA-seq tracks), and GSE47033 (mouse E10.5 RNA-seq tracks).

Acknowledgments

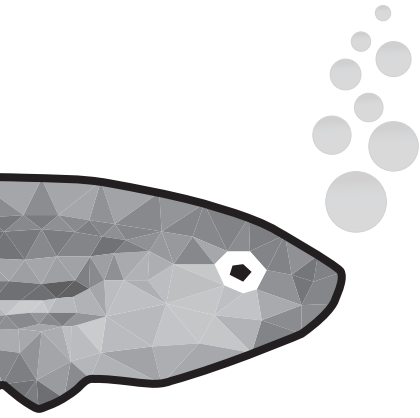
We thank Gert-Jan Veenstra and Simon van Heeringen for their critical input, Rocío Polvillo and María Nicolás-Pérez for their excellent technical help, and Iwanka Kozarewa and Lina Chen for their help on the mouse RNA-seq. The Andalusian government (JA) supported A.F.-M. as scientific manager of the Aquatic Vertebrates Platform at CABD. J.W.C. was supported by a studentship from The Institute of Cancer Research. Spanish and Andalusian government grants BFU2010-14839, CSD2007-00008, and P08-CVI-3488 to J.L.G.-S.; and BFU2011-22916 and P11-CVI-7256 to J.R.M.-M. supported this work.

References

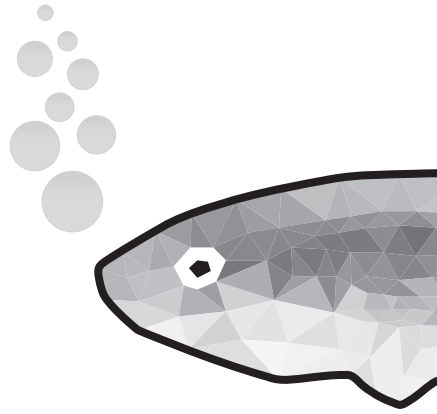
- Aday AW, Zhu LJ, Lakshmanan A, Wang J, Lawson ND. 2011. Identification of *cis* regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Dev Biol* **357**: 450–462.
- Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188**: 799–808.
- Bininda-Emonds OR, Jeffery JE, Sanchez-Villagra MR, Hanken J, Colbert M, Pieau C, Selwood L, Ten Cate C, Raynaud A, Osabutey CK, et al. 2007. Forelimb-hindlimb developmental timing changes across tetrapod phylogeny. *BMC Evol Biol* **7**: 182.
- Bogdanovic O, Fernandez-Minan A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJ, Gomez-Skarmeta JL. 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res* **22**: 2043–2053.
- Bogdanovic O, Fernandez-Minan A, Tena JJ, de la Calle-Mustienes E, Gomez-Skarmeta JL. 2013. The developmental epigenomics toolbox: ChIP-seq and MethylCap-seq profiling of early zebrafish embryos. *Methods* **62**: 207–215.
- Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**: 825–837.
- Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25–36.
- Choudhuri A, Maitra U, Evans T. 2013. Translation initiation factor eIF3h targets specific transcripts to polysomes during embryogenesis. *Proc Natl Acad Sci* **110**: 9818–9823.
- Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B. 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* **21**: 1146–1151.
- Clarke SL, VanderMeer JE, Wenger AM, Schaar BT, Ahituv N, Bejerano G. 2012. Human developmental enhancers conserved between deuterostomes and protostomes. *PLoS Genet* **8**: e1002852.
- Collins JE, White S, Searle SM, Stemple DL. 2012. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res* **22**: 2067–2078.
- Comte A, Roux J, Robinson-Rechavi M. 2010. Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evol Dev* **12**: 144–156.
- Cotney J, Leng J, Yin J, Reilly SK, Demare LE, Emera D, Ayoub AE, Rakic P, Noonan JP. 2013. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* **154**: 185–196.
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107**: 21931–21936.
- Davidson EH. 2006. *The regulatory genome: gene regulatory networks in development and evolution*. Academic Press, Amsterdam, Netherlands.
- Davidson EH, Erwin DH. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* **311**: 796–800.
- Domazet-Lošo T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**: 815–818.
- Duboule D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev Suppl* **1994**: 135–142.
- The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799–816.
- The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**: 43–49.
- Field HA, Ober EA, Roeser T, Stainier DY. 2003. Formation of the digestive system in zebrafish. I. Liver morphogenesis. *Dev Biol* **253**: 279–290.
- Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* **312**: 276–279.
- Flicke P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. 2013. Ensembl 2013. *Nucleic Acids Res* **41**: D48–D55.
- Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**: 490–493.
- Furutani-Seiki M, Wittbrodt J. 2004. Medaka and zebrafish, an evolutionary twin study. *Mech Dev* **121**: 629–637.
- Galis F, Metz JA. 2001. Testing the vulnerability of the phylotypic stage: on modularity and evolutionary conservation. *J Exp Zool* **291**: 195–204.
- Goke J, Jung M, Behrens S, Chavez L, O'Keefe S, Timmermann B, Lehrach H, Adjaye J, Vingron M. 2011. Combinatorial binding in human and mouse embryonic stem cells identifies conserved enhancers active in early embryonic development. *PLoS Comput Biol* **7**: e1002304.
- Gould SJ. 1977. *Ontogeny and phylogeny*. Belknap Press of Harvard University Press, Cambridge, MA.
- He Q, Bardet AF, Patton B, Purvis J, Johnston J, Paulson A, Gogol M, Stark A, Zeitlinger J. 2011. High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nat Genet* **43**: 414–420.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LE, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–112.
- Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol* **59**: 190–203.
- Hong JW, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty. *Science* **321**: 1314.
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**: 498–503.
- Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.
- Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* **2**: 248.
- Iwamatsu T. 2004. Stages of normal development in the medaka *Oryzias latipes*. *Mech Dev* **121**: 605–618.

- Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**: 811–814.
- Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**: 714–719.
- Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**: 253–310.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Lee AP, Kerk SY, Tan YY, Brenner S, Venkatesh B. 2011. Ancient vertebrate conserved noncoding elements have been evolving rapidly in teleost fishes. *Mol Biol Evol* **28**: 1205–1215.
- Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**: 1725–1735.
- Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev Cell* **22**: 1101–1108.
- McEwen GK, Goode DK, Parker HJ, Woolfe A, Callaway H, Elgar G. 2009. Early evolution of conserved regulatory sequences associated with development in vertebrates. *PLoS Genet* **5**: e1000762.
- Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* **27**: 937–945.
- Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41**: D64–D69.
- Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* **8**: 1551–1566.
- Nelson AC, Wardle FC. 2013. Conserved non-coding elements and *cis* regulation: actions speak louder than words. *Development* **140**: 1385–1395.
- Ong CT, Corces VG. 2012. Enhancers: emerging roles in cell fate specification. *EMBO Rep* **13**: 423–430.
- Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A, et al. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* **22**: 577–591.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.
- Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature* **490**: 98–101.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–283.
- Raff RA. 1996. *The shape of life: genes, development, and the evolution of animal form*. University of Chicago Press, Chicago, IL.
- Richardson MK, Allen SP, Wright GM, Raynaud A, Hanken J. 1998. Somite number and vertebrate evolution. *Development* **125**: 151–160.
- Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**: 321–332.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Roux J, Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. *PLoS Genet* **4**: e1000311.
- Royo JL, Maeso I, Irimia M, Gao F, Peter IS, Lopes CS, D'Aniello S, Casares F, Davidson EH, García-Fernández J, Gómez-Skarmeta JL. 2011. Transphyletic conservation of developmental regulatory state in animal evolution. *Proc Natl Acad Sci* **108**: 14186–14191.
- Sakamoto K, Onimaru K, Munakata K, Suda N, Tamura M, Ochi H, Tanaka M. 2009. Heterochronic shift in *Hox*-mediated activation of sonic hedgehog leads to morphological changes during fin development. *PLoS ONE* **4**: e5121.
- Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**: 1036–1040.
- Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, et al. 2012. A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**: 116–120.
- Slack JM, Holland PW, Graham CF. 1993. The zootype and the phylotypic stage. *Nature* **361**: 490–492.
- Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, et al. 2014. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* **507**: 371–375.
- Sprague J, Bayraktaroglu L, Clements D, Conlin T, Fashena D, Frazer K, Haendel M, Howe DG, Mani P, Ramachandran S, et al. 2006. The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res* **34**: D581–D585.
- Taher L, McGaughey DM, Maragh S, Aneas I, Bessling SL, Miller W, Nobrega MA, McCallion AS, Ovcharenko I. 2011. Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Res* **21**: 1139–1149.
- Tan MH, Au KF, Yablonovitch AL, Wills AE, Chuang J, Baker JC, Wong WH, Li JB. 2013. RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome Res* **23**: 201–216.
- Taylor JS, Van de Peer Y, Braasch I, Meyer A. 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* **356**: 1661–1679.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.
- Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, et al. 2006. Ancient noncoding elements conserved in the human genome. *Science* **314**: 1892.
- Visel A, Minovitsky S, Dubchak I, Pennacchio LA. 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res* **35**: D88–D92.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**: 897–903.
- Wang J, Lee AP, Kodzius R, Brenner S, Venkatesh B. 2009. Large number of ultraconserved elements were already present in the jawed vertebrate ancestor. *Mol Biol Evol* **26**: 487–490.
- Wang L, Wang S, Li W. 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**: 2184–2185.
- Watanabe T, Asaka S, Kitagawa D, Saito K, Kurashige R, Sasado T, Morinaga C, Suwa H, Niwa K, Henrich T, et al. 2004. Mutations affecting liver development and function in Medaka, *Oryzias latipes*, screened by multiple criteria. *Mech Dev* **121**: 791–802.
- Woo YH, Li WH. 2012. Evolutionary conservation of histone modifications in mammals. *Mol Biol Evol* **29**: 1757–1767.
- Xiao S, Xie D, Cao X, Yu P, Xing X, Chen CC, Musselman M, Xie M, West FD, Lewin HA, et al. 2012. Comparative epigenomic annotation of regulatory DNA. *Cell* **149**: 1381–1392.
- Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, et al. 2013. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**: 1134–1148.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Received July 23, 2013; accepted in revised form April 3, 2014.



Agradecimientos



A finales de octubre de 2013, acababa de asistir al acto de defensa de la tesis doctoral de Inês Gago Rodrigues, la primera en conseguirlo en el laboratorio del Doctor Martínez Morales. Yo acababa de comenzar mi proyecto de Máster y apenas era un crío inexperto pero con mucha ilusión. Allí estaba yo, preguntándome si lograría alguna vez ponerme en su piel. Recuerdo que Juan Ramón me aseguró que con esfuerzo y ganas podría conseguir lo mismo. No se equivocaba. Han pasado cinco años y ahora me veo en la misma situación que yo atisbaba con algo de envidia sana. Durante todo este tiempo, he tenido luces y sombras, he pasado por momentos de dudas y agobios, pero también de grandes alegrías que siempre compensan todo lo malo. Los resultados y conclusiones presentados en esta tesis son tan sólo la punta del iceberg de años de duro trabajo, que muestran el esfuerzo llevado a cabo a lo largo de todo este tiempo pero que creo que no reflejan los sinsabores por los que todos hemos pasado por esta etapa de nuestra vida. Sin embargo, jamás estuve solo. Mucha gente me brindó su apoyo incondicionalmente, tanto en las duras como en las maduras. A todos ellos, GRACIAS.

Gracias a mi equipo. Gracias a mi director de tesis, mi mentor, **Juan Ramón**, por ayudarme a desarrollarme profesionalmente y velar siempre por mi futuro. Todos los que me conocen saben que tiendo a ser pesimista por naturaleza, pero de Juan he aprendido que nunca hay que dar nada por perdido. De él me llevo también su capacidad de esfuerzo y su visión como científico, que han sido decisivas para modelar la mía. También tengo que darle las gracias a **Ro**, que ha sido casi una hermana mayor y una persona de total y absoluta confianza para mí. Ha estado siempre al pie del cañón y me ha sacado las castañas del fuego en muchas ocasiones, sin mostrar nunca la más mínima queja (menos por algún tirón de orejas más que merecido). Gracias a **Joaquín** por todo lo que me ha enseñado. Buena parte de lo que soy como profesional se la debo a él, y aunque haya decidido volar solo, sé que seguirá aportando su granito de arena a la ciencia y al conocimiento. Los trabajadores incansables como él “tienen el duro para cambiarlo”, como decimos por aquí. También podría decir lo mismo de **Lorena**, que aunque lleve algo menos de tiempo que yo en el laboratorio, demuestra día tras día una madurez digna de elogio, y siempre ha estado ahí para darme ánimos o para echarme un cable sin pedir nada a cambio. No me cabe duda de que tú también terminarás tu tesis y que si decides continuar tu carrera llegarás lejos. También quiero agradecer el compromiso de **María Almuedo** y **Estefanía** que, aunque lleven relativamente poco tiempo aquí, han sido decisivas para que pudiera rematar los últimos experimentos y me han dado ánimos para el último arreón. Quisiera mencionar también a otros miembros del laboratorio que ya no están aquí. Gracias a **Undu**, a **Ozren** y a **Mariana**, doctores de quien también he llegado a aprender

mucho. Gracias a **Inês** y a **María Nicolás**, con las que trabajé en mi etapa como alumno interno, que siempre me trataron como a uno más a pesar de que mi poca experiencia pudiera ser un hándicap para ellas en algunas ocasiones. Y gracias también a **Rosa**, de quien aprendí mis primeras palabras en italiano, y a **Franz**, el alemán menos alemán que he conocido en mi vida.

Un momento clave de mi periodo predoctoral fue mi estancia en Heidelberg. Allí aprendí a vivir de manera independiente y a madurar más como persona. Gracias a **Jochen** por recibirme con los brazos abiertos y a ayudar a integrarme en su equipo. Gracias a **Juan Mateo** por ser mi guía en uno de los laboratorios de investigación en teleósteos más importantes del mundo, mostrando en todo momento una paciencia infinita con mis continuas preguntas y un enorme aplomo para introducirme en el siempre complicado mundo de la Bioinformática. Gracias también a **Arturo**, que siempre está ahí para animarte aunque las cosas no salgan y al que también debo su cuota de responsabilidad en el desarrollo de esta tesis. Allí también conocí a grandes investigadores y a excelentes personas, como **Daigo**, **Lázaro**, **Jakob** y **Ali**.

Tampoco me olvido de las chicas de la plataforma. Gracias a **Ana Miñán**, la persona que me enseñó a inyectar y que siempre resolvió mis dudas. Gracias a **Ro Morales**, a **Sol**, a **Ana Álvarez**, a **Helena** y a todos los que han pasado por la plataforma, sin cuya labor, cuidando de nuestros peces, este trabajo no habría sido posible. También hay hueco para el resto de personal de servicios del centro, especialmente a **Kathy**, que me ayudó a perderle el miedo (que no el respeto) a los caros microscopios del centro, y a **Laura Tomás**, por su incansable trabajo al frente de Proteómica. Y tampoco me olvido de Administración, en especial de **Carmen**, que siempre tiene una respuesta tranquilizadora en temas burocráticos.

Gracias a **José Luis**, que aunque alguna vez hayamos tenido nuestros más y nuestros menos, ya se sabe que el roce hace el cariño y siempre se agradece la opinión sincera de un investigador de primera fila como él. Y gracias también a todos los miembros de su laboratorio, a **Juan Tena**, **Silvia**, **Nacho**, **Eli**, **Ana Neto**, **Panos**, **Rafa**, **Marta**, **Sandra**... Todos me han aportado algo, ya fuese ayuda directa o ánimos en los momentos duros.

Gracias también a mis dos tutores, **Rafa Daga** y **James**, por su apoyo y sus sugerencias después de los siempre complicados *CABD meetings* y, en general, a todas aquellas personas que me han demostrado su amistad y cariño a lo largo de este tiempo, ya fueran miembros del centro o viejos amigos: **Ann**, **Ibai**, **Sergio**, **Blanca**, **Celia**, **Juan Carlos**, **Damien**, **Isabel**, **Nico**, **Andrea**, **Alfonso**, **More**, **Miguel**, **Fran**...

Quiero reservar el final para los míos. Gracias a **Rocío**, que ya lleva más de 12 años como mi compañera de viaje; pronto habré pasado media vida con ella. Espero que no cambies nunca. Y, por último, gracias también a mi hermano, **José Ignacio**, y a **mis padres**, por aguantar mis agobios día tras día, por estar siempre ahí para preguntarme qué tal me había ido la jornada, o a llevarme en coche un sábado o un día de fiesta para terminar “esa pequeña tarea que me faltaba por hacer”. Gracias por tanto.

Yap y Taz son dos proteínas parálogas que pertenecen a una familia de activadores transcripcionales implicada en el correcto mantenimiento de tejidos adultos y en el proceso de embriogénesis, al coordinar procesos de proliferación, diferenciación celular y apoptosis mediante su unión en el núcleo a diversos factores de transcripción, fundamentalmente los pertenecientes a la familia Tead. Su actividad está controlada por tres mecanismos de regulación principales: la ruta Hippo, compuesta por una serie de kinasas que bloquean el ingreso de estos efectores al núcleo; la tensión morfogenética, que favorece el acceso de Yap y Taz al núcleo mediante un proceso estrictamente mecánico dependiente de la activación de integrinas y la polimerización de fibras de estrés; y la acción de kinasas de la familia Src/Yes (SFKs), que activan a estos parálogos mediante procesos de fosforilación dependientes e independientes de la ruta Hippo. El propósito principal de este trabajo consiste en describir la historia evolutiva de las proteínas de la familia Yap en vertebrados y analizar su papel morfogenético durante la gastrulación de teleósteos. El desarrollo de la presente tesis doctoral ha permitido demostrar la existencia de un tercer parólogo de la familia Yap, bautizado como Yap1b, que presenta una estructura divergente de Yap y Taz, con un dominio de activación transcripcional C-terminal que carece de residuos de tirosina y que por tanto escapa al control ejercido por kinasas SFK. El gen yap1b, que cuenta con una importante contribución materna pero cuyo nivel de expresión desciende súbitamente tras la activación del genoma cigótico, mantiene una relación epistásica negativa con el gen yap1, puesto que su mutación simultánea agrava el fenotipo de embriones mutantes yap1^{-/-} en medaka. La combinación de experimentos de iDamID-seq, RNA-seq, inmunofluorescencia, hibridación in situ e imaging demuestra que Yap desempeña un papel destacado en el establecimiento de los ejes principales del embrión de medaka y en el proceso de convergencia-extensión de células precursoras durante la gastrulación al promover la expresión de genes implicados en procesos de adhesión y migración celular, fundamentalmente genes que codifican para proteínas de la familia CCN como ctgfa y cyr61.